

MASTEROPPGAVE

**Masterstudium i skolerettet utdanningsvitenskap med fordypning
i matematikdidaktikk**

Mai 2021

Relativ alderseffekt og mestringsforventninger i matematikk

En kvantitativ studie av sammenhengen mellom relativ alderseffekt og mestringsforventninger i matematikk hos elever på 9.trinn i Norge

Anna Iversen

OSLOMET

OsloMet – storbyuniversitetet

Fakultet for lærerutdanning og internasjonale studier

Institutt for grunnskole- og faglærer utdanning

Sammendrag

Forskning viser at det er sammenheng mellom alder og prestasjon i matematikk i et og samme skolekull. De som er født tidlig på året ser ut til å ha et fortrinn i forhold til de som er født sent på året. Samtidig viser forskning at mestringsforventninger har positiv effekt på akademisk prestasjon og engasjement, innsats og målsetninger. Hensikten med denne studien er å sammenfatte disse to forskningsfeltene, og undersøke om det er sammenheng mellom når man er født på året og mestringsforventninger i matematikk. For å kunne gjøre dette har jeg sammen med to medstudenter og veileder utviklet og validert et instrument som måler mestringsforventninger i matematikk hos elever på 9.trinn i Norge. I tillegg til å undersøke alle elever født henholdsvis tidlig og sent i det akademiske året, undersøkes det også om det er kjønnsforskjeller, forskjeller mellom høyt- og lavt-presterende elever og om elevene rapporterer ulik grad av mestringsforventning avhengig av emneområdene i matematikk slik de er delt inn i The Trends in International Mathematics and Science Study (TIMSS).

Undersøkelsen baserer seg på et kvantitativt datamateriale som er samlet inn i forbindelse med forskningsprosjektet SESAM (*Self-efficacy – its sources and achievement in mathematics*). Utvalget består av 583 elever på 9.trinn fra 13 skoler, fra fire ulike fylker; Møre og Romsdal, Oslo, Viken og Agder. Studien har til hensikt å belyse følgende problemstilling: *Er det sammenheng mellom fødselsmåned og mestringsforventninger i matematikk for elever på 9.trinn? Eventuelt hvilken sammenheng er det? Og hvordan kan dette måles?*

For å kunne besvare problemstillingen er det utarbeidet fem forskningsspørsmål, samt hypoteser knyttet til hvert, basert på teori og tidligere forskning. Utvikling av måleinstrumentet er viet et eget forskningsspørsmål, og valideringen av dette instrument får dermed endel plass i oppgaven. Rasch-analyse er benyttet i denne prosessen, og resultatene viser at instrumentet fungerer etter sin hensikt. De øvrige forskningsresultatene viser at relativt eldre elever har høyere mestringsforventninger enn sine relativt yngre medelever. Videre viser resultatene at det ikke er sammenheng mellom RAE og mestringsforventninger knyttet til emneområdene tall, algebra, geometri og statistikk. Det ble heller ikke funnet forskjeller i forventet mestring mellom jenter og gutter. Det er signifikant forskjell i mestringsforventninger mellom lavt-presterende elever, i den retningen at relativt eldre lavt-

presterende elever har høyer mestringsforventninger. Denne forskjellen ble ikke funnet blant de høyt-presterende elevene født i ulike kvartal.

Abstract

Research shows that there is a relation between age and achievement in mathematics for students in the same school cohort. Being amongs the oldest students in a cohort can lead to persistent benefits. At the same time, research shows that self-efficacy beliefs have a positive effect on academic achievement, commitment and efforts. The purpose of this study is to summarize these two research fields, and investigate whether there is a relation between students' birth month and their self-efficacy beliefs in mathematics. In order to do this, the research group SESAM (Self-efficacy - its sources and achievement in mathematics) has developed and validated an instrument measuring self-efficacy beliefs in mathematics for students in grade 9 in Norway. I investigate differences in mathematics self-efficacy in students born early and late in the academic year, between genders, and between high- and low-performing students. Additionally, I investigate if students report different degrees of self-efficacy beliefs in different subject areas in mathematics as they appear in The Trends in International Mathematics and Science Study (TIMSS).

This study is based on a quantitative data material collected by the research group SESAM. The sample consists of 583 students in grade 9 from 13 schools, from four different counties; Møre og Romsdal, Oslo, Viken and Agder. The overall research question for the thesis is: *Is there a relation between month of birth and self-efficacy beliefs in mathematics for students in grade 9? If so, what relation is it? Moreover, how can we measure such self-efficacy beliefs?*

In order to answer the overall research question, five research questions have been prepared, as well as hypotheses related to each question. These are based on theory and previous research. The development and validation process of the mathematics self-efficacy instrument is given a central role in this master thesis. Rasch analysis is used both in the validation process and in analyzing the data. I find significant correlations between birth month and mathematics self-efficacy beliefs. However, when divided into subject areas, no significant correlation is found. The same goes for gender. There is a significant correlation in self-efficacy beliefs in low-performing students, where the relatively older low-performing students have higher self-efficacy beliefs. This difference was not found amongst the high-achieving students born in different quarters.

Forord

En langvarig, men svært lærerik prosess er forbi og jeg er både stolt og lettet over å ha fullført masterstudiet ved OsloMet. Jeg ser frem til å legge fra meg masteroppgaven og studenttilværelsen, og ta fatt på nye utfordringer som vil møte meg i hverdagen som lektor i matematikk. Det er mange som har bidratt og hjulpet meg det siste året. En stor takk rettes til min veileder Annette Hessen Bjerke som alltid har vært tilgjengelig for uendelig mange spørsmål, for oppmuntrende ord og tilbakemeldinger. Du har hatt enorm betydning gjennom hele prosessen.

Takk til medstudentene jeg har delt lesesal med for gode samtaler og trivelige pauser. Så en stor takk til deg, Elisabeth. For alt de siste fem årene, alle eksamensperioder og ikke minst dette året. Jeg er glad vi har fullført dette sammen. Til slutt rettes en spesiell takk til Sander, for at du alltid har tro på meg og oppmuntret meg til å fortsette når arbeidet har opplevdes vanskelig.

Anna Iversen
Oslo, mai 2021

Innholdsfortegnelse

SAMMENDRAG	II
ABSTRACT.....	IV
FORORD	V
LISTE OVER TABELLER.....	IX
LISTE OVER FIGURER	IX
FORKORTELSER	X
1. INNLEDNING	1
1.1 <i>Begrunnelse for valgt av tema.....</i>	<i>1</i>
1.2 <i>Forskningsprosjektet SESAM.....</i>	<i>3</i>
1.3 <i>Disposisjon av oppgaven.....</i>	<i>3</i>
1.4 <i>Problemstilling og formål med studien</i>	<i>3</i>
2. TEORI	6
2.1 MESTRING.....	6
2.2.1 <i>Mestring og kjønn</i>	<i>8</i>
2.2 MESTRINGSFORVENTNINGER (SELF-EFFICACY)	9
2.2.1 <i>Banduras teori om mestringsforventninger</i>	<i>9</i>
2.2.2 <i>Mestringsforventninger og prestasjon i matematikk.....</i>	<i>12</i>
2.2.3 <i>Mestringsforventninger og kjønn</i>	<i>15</i>
2.3 RAE - RELATIV ALDERSEFFEKT.....	16
2.3.1 <i>Biologibasert teori</i>	<i>17</i>
2.3.2 <i>Psykososial teori</i>	<i>18</i>
2.3.3 <i>Relativ alderseffekt og prestasjon i matematikk.....</i>	<i>20</i>
2.3.4 <i>Relativ alderseffekt og kjønn.....</i>	<i>21</i>
2.3.5 <i>Langtidseffektene av relativ alderseffekt.....</i>	<i>23</i>
2.4 FORSKNINGSPØRSMÅL OG FORVENTENDE SAMMENHENGER	24
3. METODE.....	26
3.1 METODISK TILNÆRMING	27
3.1.1 <i>Spørreundersøkelse som forskningsmetode</i>	<i>28</i>
3.2 RASCH RATING SCALE MODEL.....	30
3.2.1 <i>Rasch-modellen og logits</i>	<i>31</i>
.....	33
3.2.2 <i>Invarians og endimensjonalitet</i>	<i>34</i>
3.3.3 <i>Fit</i>	<i>35</i>
3.3 VALIDITET OG RELIABILITET.....	37
3.4 UTVIKLING AV MÅLEINSTRUMENTENE	39
3.4.1 <i>Bakgrunnsvariabler.....</i>	<i>40</i>

3.4.2 Mestringsforventninger	40
3.4.2 Kilder til mestringsforventninger	41
3.4.5 Mestring i matematikk	42
3.5 Pilot-studie	42
3.5.1 Analyse og resultat	43
3.6 HOVEDUNDERSØKELSEN	44
3.6.1 Populasjon og utvalg	45
3.6.2 Feilkilder	47
3.7 STATISTISKE ANALYSER	48
3.7.3 Gjennomsnittsanalyse	48
3.8 ETIKK OG PERSONVERN	48
3.8.1 Søknad til Norsk senter for Forskningsdata (NSD)	49
3.8.2 Samtykke	49
4. RESULTATER	50
4.1 VALIDERING AV MÅLEINSTRUMENTET	50
4.1.1 Instrumentets reliabilitet	51
4.1.2 Endimensjonalitet	51
4.1.3 Invarians	54
4.1.4 Oppsummering av valideringsprosessen	55
4.2 ØVRIGE FORSKNINGSPØRSMÅL	55
4.2.1 Relativ alder og mestringsforventninger	56
4.2.2 Mestringsforventninger, RAE og kjønn	61
4.2.3 Høyt-presterende- og lavt-presterende elever, mestringsforventninger og relativ alder	65
5. DISKUSJON	70
5.1 ET INSTRUMENT – MESTRINGSFORVENTNINGER I MATEMATIKK	70
5.2 RELATIV ALDER OG MESTRINGSFORVENTNINGER I MATEMATIKK	73
5.3 RELATIV ALDER OG MESTRINGSFORVENTNING KNYTTET TIL DE FIRE EMNEOMRÅDENE FRA TIMSS	79
5.4 RELATIV ALDER OG MESTRINGSFORVENTNINGER KNYTTET TIL KJØNN	80
5.5 RELATIV ALDER OG MESTRINGSFORVENTNINGER FOR HØYT-OG LAVT-PRESTERENDE ELEVER	83
5.6 PRAKTISKE IMPLIKASJONER	88
5.7 METODISKE BETRAKTNINGER	89
6. AVSLUTNING	93
6.2 VIDERE FORSKNING	94
LITTERATURLISTE	96
VEDLEGG 1 - HOVEDUNDERSØKELSEN	102
VEDLEGG 2 – RETNINGSLINJER FOR GJENNOMFØRINGEN	106
VEDLEGG 3 – INFORMASJONSSKRIV TIL FORESATTE	107

VEDLEGG 4 – PILOTUNDERSØKELSEN	108
VEDLEGG 5 – TILBAKEMELDING FRA NSD	114

Liste over tabeller

Tabell 1 - Fit-statistikk for hele utvalget (N=583)	52
Tabell 2 - Kategoristruktur basert på hele utvalget (N=583). Her vist for item 2.....	53
Tabell 3 - Verdier for Cohens d. Kilde: Cohen et al. (2018, s. 746)	56
Tabell 4 - Mestringsforventninger for elever født i 1. og 3.kvartal	57
Tabell 5 - Rangering av oppfattet vanskelighetsgrad knyttet til de ulike items. Item merket blå er fra emneområdet tall, gul=algebra, rød=geometri og grønn=statistikk.....	61
Tabell 6 - Mestringsforventninger knyttet til kjønn født i samme kvartal	62
Tabell 7 - Mestringsforventninger knyttet til kjønn født i ulike kvartal	64
Tabell 8 - Fordeling av høyt-presterende og lavt-presterende jenter og gutter født i 1. og 3.kvartal.....	65
Tabell 9 – Mestringsforventninger til høyt-presterende og lavt-presterende elever	66
Tabell 10 – Mestringsforventninger til høyt-presterende elever født i ulike kvartal	68
Tabell 11 – Mestringsforventninger til lavt-presterende elever født i ulike kvartal	68

Liste over figurer

Figur 1- Illustrasjon av variabelkartet med de ideelle linjene. Illustrasjonen viser god spredning av items og personer.	33
Figur 2 - Item Characteristics Curves for item 1, 3, 4 og 12	54
Figur 3 - Differential Test Functioning for elever født i 1.og 3.kvartal, og for jenter og gutter.	55
Figur 4 - Gjennomsnitt og spredning i mestringsforventninger for elever født i 1. og 3. kvartal	58
Figur 5 - Variabelkart som viser fordelingen av items og elever født i 1. og 3.kvartal	59
Figur 6 - Gjennomsnitt og spredning i mestringsforventninger mellom gutter og jenter født i 3.kvartal.....	63
Figur 7 - Gjennomsnitt og spredning i mestringsforventninger mellom høyt-presterende og lavt-presterende elever	67
Figur 8 - Gjennomsnitt og spredning mellom lavt-presterende elever født i 1. og 3.kvartal...	69

Forkortelser

DIF	Differential Item Functioning
DTF	Differential Test Functioning
ICC	Item Characterictics Curves
MNSQ	Mean Square
PISA	Programme for International Student Assessment
RAE	Relativ alderseffekt
RSM	Rating Scale Model
SESAM	Self-efficacy – its sources and achievement in mathematics
TIMSS	Trends in International Mathematics and Sience Study
ZSTD	Z-standardized

1. Innledning

Resultater fra Trends in International Mathematics and Science Study (TIMSS) 2015 viser at elevers motivasjon i matematikk synker jo lengre opp i grunnskolen de befinner seg (Kaarstein & Nilsen, 2016). I tillegg viser undersøkelser at et stort antall elever assosierer matematikkfaget med nederlag og mangel på interesse (Rosenlund & Gulaker, 2018). Motivasjon i matematikkfaget har fanget min oppmerksomhet mer enn noe annet.

Som nyutdannet lektor i matematikk er det viktig for meg å ha kunnskap om hvordan jeg kan legge forholdene til rette slik at mine fremtidige elever skal få mulighet til å oppleve mestring i matematikk, uavhengig av bakgrunn, kjønn, holdninger eller når på året de er født. For å ha mulighet til å gjennomføre en slik tilrettelegging er det viktig å ha kunnskap om aspekter som kan påvirke elevers prestasjoner og deres motivasjon og tro på egne evner.

1.1 Begrunnelse for valgt av tema

Jeg hadde tidlig i masterløpet sett for meg at masteroppgaven min ville ta opp tema innenfor motivasjon i matematikk, men akkurat hva jeg ønsket å undersøke hadde jeg ikke klart for meg. Høsten 2019 var jeg i en forelesning med Annette Hessen Bjerke. Hun snakket om et pågående arbeid med en forskningsartikkel som undersøkte og sammenlignet prestasjoner i matematikk hos elever født i januar opp mot elever født i desember samme år.

Undersøkelsene viste en trend om at elever født tidlig i det akademiske året presterer bedre enn sine medelever født sent samme år (Bjerke, Smestad, Eriksen & Rognes, til fagfelleevaluering). Dette omtales ofte som *relativ alderseffekt* (RAE) – nemlig konsekvensen av relativ aldersforskjell (Wattie, Cobby & Baker, 2008). Dette var mitt første møte med konseptet RAE, og tankene mine kom stadig kommet tilbake til dette resultatet. Som matematikklærer er dette viktig å være klar over!

I Norge starter alle barn født innenfor samme kalenderår samtidig på skolen. En slik regel er for foreldre enkelt å følge og medfører relativt små kostnader sammenlignet med systemer som bygges på andre skolestartprinsipper (Strøm, 2004). Det er rimelig å forvente at RAE, og konsekvensen av denne effekten, er mer fremtredende i de laveste skoletrinnene der aldersforskjellene er prosentvis større (Bjerke et al., til fagfelleevaluering). Med dette argumentet vil det også være rimelig å anta at RAE vil forsvinne jo eldre elevene blir. Likevel

viser forskningen at ulempen med å være blant de yngste i en alderskohort er tilstede også ved utgangen av grunnskolen, og ikke bare innenfor akademisk prestasjon, men også innenfor idrett (Reed, Parry & Sandercock, 2017). Noen har til og med funnet at dette kan ha konsekvenser for akkumulert lønn gjennom et yrkesliv (Solli, 2017).

Det råder mange ulike hypoteser og teorier om hvorfor RAE eksisterer. Blant annet viser forskning gjort av Gledhill, Ford og Goodman (2002) at lærere i gjennomsnitt ikke tar nok hensyn til elevenes kronologiske alder i klassen når de vurderer elevenes evner. Det viser seg også at lærere oppfatter at de relativt eldre elevene har bedre akademiske ferdigheter, og at de yngre elevene har større sannsynlighet for å bli ansett som lavt-presterende elever (Baker, Kupersmidt, Tichovolsky, Voegler-Lee & Arnold, 2015). Med andre ord er det grunn til å tro at lærere oppfatter relativt yngre elever annerledes enn de oppfatter relativt eldre elever i samme årskull. En slik selvoppfyllende profeti kan være en viktig faktor som bidrar til den varig RAE (Bjerke et al., til fagfelleevaluering). Dette kan støttes av forskning på RAE i kroppøving, som antyder at RAE kan dempes ved økt bevissthet blant lærere i skolen (Aune, Pedersen, Ingvaldsen & Dalen, 2017). Spørsmålet om dette er en selvoppfyllende profeti i seg selv, fremhever viktigheten av å undersøke disse sammenhengene mer detaljert (Bjerke et al., til fagfelleevaluering). Til tross for disse funnene er det ikke etablert noen enkel årsak eller årsakssammenheng som forklarer fenomenet RAE.

Ideen om å peile inn på motivasjon i matematikkfaget i min masteroppgave ble etter hvert farget av en økende interesse for RAE – jeg ønsker å inkludere begge forskningsfeltene i masteroppgaven min. Teorien om mestringsforventninger av Albert Bandura ble vektlagt i undervisningen i mitt masterforløp, og har en sentral posisjon innfor feltet om motivasjon og læring. Bandura argumenterte for at mestringsforventninger spiller en betydningsfull rolle i hvilke valg vi tar, og hvor mye innsats og utholdenhet vi utviser. Forskningen innenfor mestringsforventninger i matematikk (mathematics self-efficacy) bekrefter at mestringsforventninger har en direkte effekt på elevers prestasjon i matematikk.

Hvis det faktisk er slik at de eldste elevene presterer bedre enn sine yngre klassekamerater, og det samtidig er en sammenheng mellom mestringsforventninger og prestasjon, da slo det meg at det trolig også må være en sammenheng mellom når på året en er født og mestringsforventninger. Søk i flere databaser ble gjort for å finne studier som undersøker

denne koblingen, men det viste seg å være et lite utforsket felt, spesielt i Norge. Jeg bestemte meg – dette ville jeg se nærmere på.

1.2 Forskningsprosjektet SESAM

Denne masteroppgaven er en del av forskningsprosjektet SESAM - *Self-efficacy – its sources and achievement in mathematics*. Høsten 2020 har jeg og to andre masterstudenter deltatt i prosjektet, med Annette Hessen Bjerke som prosjektleder. Vi har i fellesskap utviklet og oversatt til sammen tre ulike instrumenter. Vi har samarbeidet om innhenting av data både til pilotering og til hovedundersøkelsen. Dermed baserer alle våre tre masteroppgaver seg på det samme datamateriale, der hver av oss har ulikt fokus med individuelle forskningsspørsmål og analyser.

1.3 Disposisjon av oppgaven

Oppgaven er delt inn i seks kapitler. Inneværende innledende kapittel avrundes med masteroppgavens problemstilling. Kapittel 2 starter med en redegjørelse av teorien og litteraturen som oppgaven baserer seg på. Teori knyttet til mestringsforventninger vil fungere som et teoretisk fundament for denne masteroppgaven. På slutten av teorikapittelet operasjonaliseres problemstillingen, og forskningsspørsmålene presenteres. Videre presenteres forskningsmetode og design i kapittel 3, før utviklingen av spørreskjemaet og datainnsamling beskrives. Til slutt presenteres hvilke analyser som tas i bruk. Kapittel 4 tar for seg undersøkelsens resultater, og vil være todelt. Første del omhandler validering av måleinstrumentet for mestringsforventninger og i andre del presenteres analyseresultatene av dataene som er samlet inn med dette måleinstrumentet. De viktigste funnene diskuteres i lys av teori og tidligere forskningsfunn i kapittel 5. Oppgavens siste kapittel er et avsluttende kapittel og vil fungere som en oppsummering av det arbeidet denne masteroppgaven har ledet fremt til av ny kunnskap for forskningsfeltet.

1.4 Problemstilling og formål med studien

Mye kan tyde på at RAE forklarer noe av det faglige spennet vi opplever blant elever i samme årskull. Samtidig vet vi at elevers mestringsforventninger i matematikk også i stor grad predikerer elevers prestasjoner i matematikk. Med ideen om å se nærmere på og undersøke om det er en sammenheng mellom disse forskningsfunnene, ble følgende problemstilling ble formulert for denne masteroppgaven;

Er det sammenheng mellom fødselsmåned og mestringsforventninger i matematikk for elever på 9.trinn? Eventuelt hvilken sammenheng er det? Og hvordan kan dette måles?

Ved gjennomgang av teori og tidligere forskning om RAE knyttet til mestring i skolefag og mestringsforventninger i matematikk ble jeg oppmerksom på at mange studier også inkluderer kjønnsforskjeller i sine undersøkelser. Både innenfor mestringsforventninger og RAE viser det seg å være kjønnsforskjeller, i den generelle forstand at gutter ofte har høyere mestringsforventninger enn jenter (Jensen & Nordtvedt, 2013) og at den relative alderseffekten gjennomgående er høyere for gutter (Solli, 2017). På bakgrunn av dette valgte også jeg å inkludere kjønn som en variabel i mine analyser. Dette ble gjort for å nyansere bildet og for å kunne undersøke eventuelle sammenhenger mer i dybden. Jeg har også valgt å inkludere elevenes prestasjon, for å se om en eventuell sammenheng mellom fødselsmåned og mestringsforventninger opptrer forskjellig hos høyt-presterende enn hos lavt-presterende elever. Problemstillingen og disse tilhørende utdypende momentene er dermed utgangspunktet for denne masteroppgaven. Problemstillingen blir ytterligere operasjonalisert i slutten av teoridelen i form av fem ulike forskningsspørsmål.

Masteroppgaven har ikke til hensikt å bevise kausale sammenhenger mellom fenomenene som undersøkes, men å belyse mulige relasjoner mellom fødselsmåned og mestringsforventninger i matematikk. Formålet med studien er å skaffe mer kunnskap om RAE i den norske skolen, og å belyse mulige sammenhenger med motivasjonsteorien om mestringsforventninger. Som lærer har man et ansvar for å legge til rette for læring for alle elever, og i tillegg stimulere elevens lærelyst å tro på egen mestring (Kunnskapsdepartementet, 2020). Alle elever møter skolen med ulike erfaringer, holdninger, forkunnskaper og behov, og som lærere skal du etterstrebe å gi alle likeverdige muligheter til læring og utvikling, uavhengig av hvilke forutsetninger elevene har. Forskningen viser at relativt eldre elever presterer bedre enn de yngre elevene i årskullene, og at dette er et fenomen som vedvarer gjennom hele grunnskolen (Solli, 2017). I tillegg viser forskningen at lærere oppfatter de relativt eldre elevene som mer faglig dyktig (Baker et al., 2015), og at lærere smiler mer, gir mer faglig støtte og oppmuntring til elever som forventes å prestere godt (Jussim, Robustelli & Cain, 2009). Over omtalte jeg dette som selvoppfyllende profeti.

For at lærere skal kunne gi likeverdig opplæring er det en forutsetning å vite hvilke aspekter som gjør elever forskjellige, eller som gjør at elever oppfattes forskjellig. I overordnet del i læreplanverket påpekes det at skolens forventning til innsats og mestring til hver enkelt elev, påvirker læringsmulighetene og troen på egne evner (Kunnskapsdepartementet, 2020). Hvis det er slik at relativt yngre elever presterer dårligere, har lavere mestringsforventninger og lærerne har lavere forventninger til dem, da er dette viktige aspekter alle lærere bør være bevisste. For å i det hele tatt ha mulighet til å kunne, i den grad det er mulig, motvirke disse forskjellene er det en absolutt forutsetning å ha kjennskap til denne problematikken. Det har denne masteroppgaven til hensikt å belyse og formidle.

2. Teori

For å kunne besvare den overordnede problemstilling vil jeg i dette kapittel gi leseren ett innblikk i den forskningen masteroppgaven orienterer seg i forhold til, samt oppgavens teoretiske rammeverk. Først vil jeg kort presentere forskning på mestring i matematikk, og funn knyttet til kjønnsforskjeller innen mestring. Med begrepet mestring mener jeg å belyse hvor godt elevene presterer i matematikkfaget. Videre presenteres Banduras teori om mestringsforventninger, som fungerer som et teoretisk rammeverk for oppgaven. Her presenteres også tidligere forskning på feltet, der forskning på relasjonen mellom mestringsforventninger og prestasjon i matematikk belyses spesielt. Videre legges empiri om kjønnsforskjeller knyttet til mestringsforventninger frem. I neste del av teorikapittelet blir RAE introdusert. Først fokuseres det på to teorier om hvorfor RAE eksisterer, før tidligere forskning trekkes inn. Videre presenteres forskning som peker på kjønnsforskjeller knyttet til RAE. Til slutt presenteres forskningsspørsmålene. I de tilfellene jeg har oversatt engelske fagbegreper til norsk setter jeg de engelske begrepene i parentes for at leseren lettere skal se koblingen til originalkildene som jeg har rettet arbeidet mitt mot.

2.1 Mestring

Matematikk anses som et allmenndannende fag hvor gode regneferdigheter ligger til grunn for å kunne mestre faget. I tillegg regnes det som ett av de fagene som krever mest innsats av elevene i opplæringen (Kunnskapsdepartementet, 2011). At mange elever har et dårlig forhold til og opplever faget som vanskelig, gir store utfordringer for matematikklærere. I følge Kunnskapsdepartementet (2011) er det en betydelig høyere andel elever som får karakter én eller to ved avsluttende eksamen i grunnskolen i matematikk, sammenlignet med de andre fagene. Likevel viser resultater fra TIMSS 2015 at norske elever verdsetter matematikkfaget i større grad enn før, og at de har et bedre selvbilde knyttet til matematikk enn tidligere. Resultater fra Programme for International Student Assessment (PISA) 2018 viser også at norske elever presterte over OECD-gjennomsnittet i matematikk, og presterte noe bedre i 2018 enn sist matematikk var hovedområde, i 2012 (Jensen et al., 2019). Videre viser resultater fra TIMSS-undersøkelsen viser at norske elever på 8.trinn har gjennomgått en relativ stor og positiv endring i matematikkprestasjon fra 2003 til 2015 (Olsen & Blömeke, 2018).

TIMSS rapporterer resultater om matematikkprestasjon for ungdomstrinnet ut fra de fire ulike emneområdene tall, statistikk, geometri og algebra. Det er viktig å fremheve at TIMSS vektlegger en rekke problemløsningssituasjoner innenfor matematikk, og at omtrent to tredjedeler av oppgavene krever mer enn kun memorert kunnskap (L. S. Grønmo, Lindquist, Arora & Mullis, 2013). Sammenlignet med oppgavene som gjennomføres på 5.trinn har oppgavene til elevene på 8.trinn mindre vekt på «å vite» og større vekt på resonering. Innenfor emneområdet tall på 8.trinn forventes det at elevene kan demonstrere forståelse av heltall og rasjonale tall, ved for eksempel å identifisere, sammenligne og sortere brøker og desimaltall. Innenfor emneområdet algebra skal elevene løse autentiske oppgaver ved hjelp av algebraiske modeller og forklare sammenhenger som involverer algebraiske konsepter. Det påpekes av L. S. Grønmo et al. (2013) at elevene må gå utover enkle memoreringsteknikker for å kunne forklare og forstå de algebraiske uttrykkene. Det understrekes at elevene trenger konseptuell forståelse for å ta i bruk og forstå lineære funksjoner. Emneområdet geometri legges opp på en slik måte at elevene skal kunne beskrive egenskaper til en rekke to- og tredimensjonale figurer, og vise kompetanse innenfor geometriske målinger (omkrets, areal og volum). Denne kunnskapen skal kunne tas i bruk for å løse autentiske problemer. Innenfor statistikk blir tradisjonelle diagrammer stadig mer kompliserte og erstattet av en rekke nye grafiske former. Det forventes at elevene skal kunne lese og trekke ut informasjon fra disse. Det vektlegges også at elevene er kritiske til fremstillingen av slike tabeller og diagrammer, slik at de ikke villedes av forvrengt fremstilling av data. Elevene må også vise kunnskap om statistikk, og forståelse av noen begreper knyttet til sannsynlighet.

Det som preger den norske profilen er at det er stor variasjon i hvor godt elevene mestrer oppgavene innenfor hvert av emneområdene (Bergem, 2016). Elevene skårer høyt innenfor tall og enda høyere innenfor emneområdet statistikk, men atskillig lavere i geometri og klart lavest innenfor algebra. Spesielt innenfor algebra presterer de norske elevene svakt, og dette er en åpenbar svakhet i de norske elevenes matematiske kompetanse. Den norske profilen er ikke så forskjellig fra den svenske og engelske, men den norske profilen har vesentlig større forskjeller i skår mellom de ulike emneområdene (Bergem, 2016). USA har derimot en mye flatere profil, med klart mindre forskjeller mellom emneområdene. Resultatene fra TIMSS 2015 viser at det største forskjellen mellom amerikanske og norske elever er nettopp innenfor algebra, hvor de amerikanske elevene skårer 54 poeng høyere enn de norske. En kan selvsagt stille spørsmål om hvorfor det er slik i Norge, og hva som gjør det så annerledes i USA.

Denne oppgaven forfølger ikke dette og lignende spørsmål, men den norske profilen tas med som et viktig bakteppe.

Selv om norske elevers prestasjon på standardiserte matematikkprøver har økt de siste årene, og vi med dette ser konturene av en positiv trend, er det ifølge Jensen et al. (2019) fortsatt én av fem elever som presterer på lavt nivå i norske ungdomsskoler. TIMSS 2019 viser at det er en signifikant nedgang i emneområdene tall og statistikk for elevene på ungdomstrinnet, mens prestasjonen i algebra og geometri er stabil (Kaarstein, Radišić, Lehre, Nilsen & Bergem, 2020). Med dette ser det ut til at den positive trenden flater litt ut. Det var også en økning med fem prosentpoeng i antall lavt-presterende elever på ungdomstrinnet. Dette er tilsvarende andel som i Sverige og Island, mens den er noe lavere enn for Danmark og Finland. Til tross for at resultatene fra TIMSS og PISA viser positiv gjennomsnittlig endring for norske elever, så er det fremdeles forskjeller i mestring mellom jenter og gutter (Zhu, 2007).

2.2.1 Mestring og kjønn

En gjennomgang av forskningslitteraturen gjort av Zhu (2007) i tiårene rundt årtusenskiftet viste at flere studier rapportert om kjønnsforskjeller i prestasjon på standardiserte tester i matematikk, i gutters favør. Gallagher og Kaufman (2005) påpeker at mye forskning innen kjønn og matematikk fokuserer på forskjeller mellom kjønnene, til tross for at det i mange tilfeller er flere likheter. De påpeker også at forskjellene innad i kjønnene sannsynligvis er større enn forskjellene mellom kjønnene. Til tross for dette finnes det tydelige kjønnsforskjeller på alle nivåer i utdanningssystemet i Norge (NOU 2019: 3, 2019), der det i grunnskolen viser seg å være størst kjønnsforskjeller i lesing – denne gangen i jenters favør (Backe-Hansen, Walhovd & Huang, 2014).

I matematikk har trenden vært det motsatte i den betydning at gutter har prestert bedre enn jenter på nasjonale prøver i regning både på 5., 8. og 9.trinn (Statistisk sentralbyrå, 2017). Imidlertid viste resultatene fra PISA 2018 at for første gang i PISAs historie presterer norske jenter bedre enn norske gutter i alle de tre fagområdene lesning, matematikk og naturfag (Jensen et al., 2019). Kjønnsforskjellene er store innenfor lesing, men gjennomgående små for matematikk og naturfag. Det påpekes imidlertid videre at for andelen elever som regnes som lavt-presterende i matematikk, er gutter i større grad representert. Jentene oppnår også bedre resultater både på eksamen på 10.trinn og som standpunkt karakter i matematikk (NOU 2019: 3, 2019). Paradoksalt nok, og til tross for at jenter i tillegg oppnår i gjennomsnitt flere

grunnskolepoeng, og er overrepresentert i høyere utdanning, er det få jenter som studerer matematikk videre etter fullført grunnskole (Foyn, 2019).

Allerede fra tidlig barndom viser jenter mer utviklede ferdigheter innenfor selvregulering og planmessighet, samt emosjonelle- og sosiale ferdigheter enn gutter (NOU 2019: 3, 2019). Guttene utvikler generelle kognitive evner på senere stadier enn jentene. Tidlig utvikling av kognitive evner kan ha betydning for hvor mye kunnskap et barn kan tilegne seg før skolestart. En slik tidlig utvikling vil påvirke videre læring i skolen, slik at det blir lettere å lære på et senere tidspunkt i skoleløpet (NOU 2019: 3, 2019). Det er også atskillig flere gutter enn jenter som får diagnoser som ADHD (Crawford, Dearden & Greaves, 2013b; Gledhill et al., 2002; Karlstad, Furu, Stoltenberg, Håberg & Bakken, 2017; NOU 2019: 3, 2019), språkforstyrrelser, autismespekterforstyrrelser og andre utviklingsforstyrrelser som kan spille en rolle for utvikling, læring og psykisk helse (NOU 2019: 3, 2019). Av dette kan vi lese at bildet er komplekst.

2.2 Mestringsforventninger (self-efficacy)

Troen elevene har på at egne evner vil gjøre at de lykkes i en gitt situasjon, er en viktig pådriver for at elevene på et senere tidspunkt vil lykkes eller mislykkes i lignende situasjoner (Pajares, 2006). Elever blir ofte tvunget til å fullføre eller delta i aktiviteter de ikke velger selv, men så snart de får muligheten til å velge, vil de mest sannsynlig velge oppgaver eller aktiviteter de mener er oppnåelig med deres evner, og unngå de oppgavene de tror de vil mislykkes med. Mestringsforventninger spiller en avgjørende rolle for valg av aktiviteter, for innsats og for utholdenhet når ting butter imot, men virker også bestemmende for den grad av ro eller angst man opplever i møte med ulike oppgaver (Usher & Pajares, 2008). Vi tenderer til å unngå situasjoner vi tror overgår våre evner, men involverer oss og tilpasser atferden vår til aktiviteter vi tror vi kan mestre (Bandura, 1997).

2.2.1 Banduras teori om mestringsforventninger

Mestringsforventninger er et nøkkelbegrep i sosial kognitiv teori, der mestringsforventninger “refers to beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, s. 3). Med dette argumenterte Bandura (1977) for at forventninger om mestring spiller en avgjørende rolle for hvilke valg vi tar og hvor mye innsats og utholdenhet vi viser. Kognitive prosesser spiller en fremtredende rolle i det å

bevare og lære nye ferdigheter, og dermed vil mestringsforventninger spille en viktig rolle i nettopp læring. Sosial kognitiv teori består av mange av de samme tankene som fra sosial læringsteori, men anses som en mer helhetlig teori om læring, motivasjon, mestringsforventninger og regulering av atferd (Skaalvik & Skaalvik, 2018). Ifølge denne teorien skjer læring hovedsakelig gjennom egne erfaringer eller gjennom observasjon.

Sosial kognitiv teori er forankret i det Bandura (1977) betegner som *human agency*, som innebærer at mennesker handler ut fra en egen intensjon. Skaalvik og Skaalvik (2018) beskriver dette som å være «agent i eget liv». Av alle elementer som påvirker oss i å være agent i eget liv, er det ingenting som er mer sentralt enn å ha tro på egne evner og til å ha kontroll over valgene som påvirker livene våre (*causative capabilities*) (Bandura, 1989). Å være agent i eget liv handler om menneskers ønske om å ha kontroll over egne handlinger, tanker og følelser. Med andre ord; om man tror at man ikke har evnene som skal til for å oppnå ønsket resultat, er det ingen grunn til i det hele tatt å prøve (Bandura, 1977). Av dette leses viktigheten av å ha høye mestringsforventninger.

Teorien om mestringsforventninger er fremtidsorientert og retter seg mot hva en tror en kan få til. Bandura (1977) skiller mellom forventning om mestring (*efficacy expectations*) og forventet utkomme (*outcome expectations*), og mener at begge har betydning for hvilke valg en tar. Forventning om mestring omfatter de forventningene en har for å lykkes med å utføre en bestemt oppgave. Det er dette som kalles mestringsforventninger. Forventet utkomme handler om en persons forventning om at en gitt handling vil føre til et bestemt utfall, altså konsekvensene av handlingene. Det understrekes at disse begrepene er forskjellige fordi en person kan forvente at en gitt handling kan føre til et bestemt utfall, men hvis en tviler på om en kan utføre denne handlingen, vil ikke denne informasjonen påvirke valget en tar. Til tross for at både forventning om mestring og forventet utkomme er vesentlige deler av Banduras teori, vil jeg i denne masteroppgaven fokusere på forventning om mestring, og dermed teori om mestringsforventninger.

Mestringsforventninger skal ikke likestilles med selvtillit (Bandura, 1977).

Mestringsforventninger handler om vurdering av egne evner, og er ikke en vurdering av ens egen selvverd (som er noe av det som ligger i konstruktet selvtillit). Zimmerman (1995) skriver at “Student judge their capabilities to fulfill given task demands, not who they are as people or how they feel about themselves in general” (s.203). Det vil derfor være viktig å

påpeke at mestringsforventninger og selvtillit ikke omfavner det samme. Elever vurderer om de er i stand til å løse en gitt oppgave, ikke hva de føler om seg selv. En kan vurdere seg selv som håpløs ineffektiv i en gitt situasjon, uten at det trenger å påvirke elevens selvtillit (Bandura, 1977).

Forventning om mestring skapes og utvikles når elever tolker informasjon fra fire ulike kilder; *mestringserfaringer (mastery experiences)*, *vikarierende erfaringer (vicarious experiences)*, *verbal overtalelse (social persuasion)* og *psykologiske og fysiske tilstander (emotional and physiological states)* (Bandura, 1977; Wæge & Nostrati, 2018). Informasjon basert på tidligere mestringserfaringer regnes som den kilden som har mest innvirkning på mestringsforventninger. Elever vil, underveis og i etterkant av å ha fullført en oppgave, tolke og evaluere resultatet (Usher & Pajares, 2008). Hvis elevene vurderer at de har løst oppgaven med suksess, vil forventningene om å mestre en lignende oppgave eller aktivitet øke. Om en elev legger innsats i å løse en oppgave, men opplever å mislykkes, vil forventningen om mestring synke. Selv om mestringserfaringer regnes som den viktigste kilden til mestringsforventninger (Bandura, 1977), vil elever også skape og utvikle mestringsforventninger ut fra de andre tre kildene. Vikarierende erfaringer har mest verdifull informasjon om det er visse likheter mellom den som observerer og den som blir observert, som kjønn, alder og etnisitet. Likevel påpeker Bandura (1997) at slike likheter ikke nødvendigvis er avgjørende for at informasjonen skal ha verdi for utvikling av mestringsforventninger, og at mange søker rollemodeller med høyer status og kunnskap enn en selv. Den tredje kilden til mestringsforventninger er verbal overtalelse, som blir mye brukt i samfunnet siden den er lett tilgjengelig for alle (Bandura, 1977). Oppmuntringer fra foreldre og lærere, som elevene har tillitt til, kan øke deres forventninger om å mestre aktiviteten de står ovenfor. Den fjerde og siste kilden handler om en persons psykologiske og fysiske tilstand, som stress, angst, utmattelse og humør. Elever lærer å tolke følelser som indikatorer for egen kompetanse. Elever som opplever negative emosjonelle reaksjoner knyttet til spesifikke aktiviteter eller oppgaver, vil sannsynligvis tolke denne følelsen som et bevis på manglende kompetanse innenfor området og dermed ha lave forventninger om å mestre aktiviteten (Usher & Pajares, 2008).

Forventninger om mestring blir av Bandura (1977) definert som en ren kognitiv prosess, hvor en foretar en vurdering av egne evner for å oppnå et resultat. Dermed handler ikke mestringsforventninger bare om å kunne utøve kontroll over handlingene sine, men det

handler også om å påvirke tankeprosesser, motivasjon og affektive og psykologiske tilstander. Forventning om mestring er heller ikke noe globalt eller kontekstløst som kan måles ved en enkelt test, men beskrives heller som et multidimensjonalt fenomen (Bandura, 1997). Av dette følger at elevenes forventning om mestring vil variere i både *omfang*, *styrke* og *generalitet*. En oppgaves *omfang* og vanskelighetsgrad, vil påvirke elevens forventning om mestring i ulik grad. En elev kan føle seg flink i matematikk, men likevel ha lave mestringsforventninger i møte med en vanskelig matematikkoppgave eller en oppgave en ikke har erfaring med (Skaalvik & Skaalvik, 2018). På samme måte kan en elev som anser seg selv som svak i matematikk ha høye mestringsforventninger hvis han får en lettere oppgave. Mestringsforventninger vil også variere i *styrke*, i den forstand at sterke forventninger om mestring gjør at elever viser større innsats og utholdenhet i møte med utfordringer. Det er heller ikke slik at en elev med høye mestringsforventninger på ett område, har dette innenfor alle matematikkens områder. Dette handler om mestringsforventningenes generalitet. En elev kan rapportere om høye forventninger om mestring innenfor addisjon og multiplikasjon, men samtidig ha lave mestringsforventninger for å løse subtraksjon- og divisjonsoppgaver (Zimmerman & Clearly, 2006). Med dette menes det at mestringsforventninger varierer avhengig av person, situasjon og kontekst. Hvilke oppgaver elevene må utføre, hvor lang tid de har til rådighet og hvilke hjelpemidler de får ta i bruk, er faktorer som kan påvirke elevenes forventning om mestring. Tidligere forskning viser at elever med lav forventning om mestring ofte trekker seg unna oppgaver og situasjoner som krever mer kompetanse enn de opplever å være i besittelse av, mens elever med høye mestringsforventninger lettere går løs på utfordringer og viser mer innsats og utholdenhet i møte med problemer (Bandura, 1977, 1997; Bandura & Schunk, 1981; Skaalvik, Federici & Klassen, 2015; Zimmerman, 1995).

2.2.2 Mestringsforventninger og prestasjon i matematikk

Som nevnt over er teorien om mestringsforventninger kontekstspesifikt, og et område i utdanningsforskningen som har fått mye oppmerksomhet er mestringsforventninger i matematikk (Pajares, 2005). Forskningsresultatene som presenteres nedenfor er stort sett sammenfallende, og stammer fra en tidsperiode fra 1995-2017. Forskningen viser at mestringsforventninger har positiv effekt på akademisk engasjement, målsetninger, valg av oppgaver, utholdenhet og innsats, valg av strategier, motivasjon, prestasjon, og til og med for valg av karriere (Bandura, 1977, 1997; Bandura & Locke, 2003; Bong & Skaalvik, 2003; Bouffard-Bouchard, 1990; Hackett & Betz, 1989; Pajares, 2006; Pajares & Miller, 1995; Skaalvik et al., 2015; Usher & Pajares, 2008; Zimmerman, 1995, 2000; Zimmerman &

Clearly, 2006). Doménech-Betoret, Abellán-Roselló og Gómez-Artiga (2017) undersøkte sammenhengen mellom mestringsforventninger, tilfredsstillelse og prestasjon hos 797 spanske elever i alderen 12-17 år. Resultatene viste at mestringsforventninger hadde en signifikant, men indirekte effekt på elevenes akademiske prestasjon. Med dette påpekte de at analysene viste at forventninger om mestring hadde direkte effekt på forventet utkomme (expectancy-value beliefs) som igjen hadde en signifikant og direkte effekt på akademisk prestasjon. På bakgrunn av dette konkluderte de med at mestringsforventninger har en betydelig effekt på elevenes prestasjon i skolen.

Pajares og Miller (1995) gjennomførte en studie med 391 studenter, og fant at studentenes tillitt til å lykkes i å løse spesifikke matematikkoppgaver var en signifikant predikasjon for deres evner i matematikk. Lignende resultat ble gjort av Pajares og Kranzler (1995) som fant at mestringsforventninger i matematikk hadde en sterk og direkte effekt på elevenes prestasjoner innenfor problemløsning, selv når det ble kontrollert for kognitive evner. Roick og Ringeisen (2018) studerte førsteårsstudenter i høyere utdanning, og undersøkte sammenhengen mellom prestasjon i matematikk, mestringsforventninger og læringsstrategier. Resultatene viste at grad av mestringsforventninger var en god indikator til elevenes matematikkarakterer, og at høy grad av mestringsforventninger var relatert til en bedre eksamensprestasjon. I tillegg viste studien at høyere mestringsforventninger også var positivt assosiert med større bruk av både kognitive- og metakognitive læringsstrategier. Street, Malmberg og Stylianides (2017) fant også sammenheng mellom mestringsforventninger og prestasjon i matematikk i sin undersøkelse av norske elevers skår på nasjonale prøver i regning og mestringsforventning. Totalt deltok 756 elever fra 5.trinn, 8.trinn og 9.trinn, og resultatene viste at det var sammenheng mellom elevenes forventning om mestring og skår på nasjonale prøver på alle trinn. Imidlertid viste analysene at det var sterkere relasjon mellom mestringsforventninger og skår på nasjonale prøver for elevene på 8. og 9.trinn, enn det var hos elevene på 5.trinn.

Pajares og Kranzler (1995) fant i tillegg at elevene i deres utvalg overvurderte sine matematiske evner, i den grad at de ikke samsvarte med kompetansen elevene viste. Hackett og Betz (1989) fant lignende tendenser der 43% av jentene og 54% av guttene overvurderte sine evner. Å vurdere om man er i stand til å løse en oppgave eller ikke er ingen enkel sak (Bandura & Schunk, 1981). Mange elever er ikke klar over de nødvendige kognitive ferdighetene som trengs for å løse en spesifikk matematikkoppgave. I tillegg påpekes det av

Bandura og Schunk (1981) at i undersøkelser som studerer mestringsforventninger, er eksponeringstiden elevene får med hver enkel oppgave svært kort. Matematikkoppgaver kan framstå som tilsynelatende enkle, men likevel inneholde komplekse regneoperasjoner. Dette kan fremstå misvisende for mange elever. I slike tilfeller kan uoverensstemmelser mellom forventning om mestring og faktisk prestasjon stamme fra misforståelser om oppgavens vanskelighetsgrad, så vel som dårlig selvinnsikt. På bakgrunn av dette understreker Bandura og Schunk (1981) at det ikke er overraskende at elever i noen tilfeller overvurderer evnene sine, spesielt på oppgaver som fremstår som enklere enn de i realiteten er. Chen (2003) diskuterer også dette, og påpeker at de høyt-presterende elevene ofte er mer nøyaktig i sine vurderingen av egen kompetanse og evner i matematikk enn lavt-presterende elever. Hun fant også at elevene var mer nøyaktig i vurdering av egne evner for enklere oppgaver enn for vanskelige oppgaver. Dette støttes av funnene i Pajares og Graham (1999) som påpeker at de lavt-presterende elevene ofte overvurderte seg selv, og på den måten også var vesentlig mindre nøyaktig i vurderingen av egne evner. Samlet viser dette flere grunner til at det kan være vanskelig for den enkelte elev å rapportere egne mestringsforventninger.

Det er godt dokumentert at mestringsforventninger har positiv effekt på prestasjon i matematikk (Hackett & Betz, 1989; Pajares & Kranzler, 1995; Pajares & Miller, 1995; Pietsch, Walker & Chapman, 2003; Randhawa, Beamer & Lundberg, 1993; Skaalvik et al., 2015). Til tross for dette kan det virke som at elevene med høye mestringsforventninger opplever disse positive virkningene i større grad enn elever med lave mestringsforventninger, til tross for like ferdigheter. Bandura (1977) påpeker dette og skriver at mestringsforventninger ikke handler om hvor mange ferdigheter man besitter, men om hva en tror en kan få til med de ferdighetene en har. Dette påpekes også av Chen (2003) som skriver at selv om elevene rapporterer om høy forventning om mestring, vil ikke dette produsere den kompetansen elevene trenger for å prestere, om de i utgangspunktet mangler de nødvendige ferdighetene. Forskjellige elever med like ferdigheter, eller den samme personen under ulike omstendigheter, kan prestere ulikt avhengig av svingninger i forventning om mestring. Dyktighet kan lett overstyres av tvil, i den forstand at også talentfulle individer gjør dårlig bruk av ferdighetene sine under omstendigheter hvor troen på seg selv undergraves (Bandura, 1997, s. 37).

For å underbygge dette vises det til to undersøkelser. Collins (1982) studerte elever og deres evner i problemløsning i matematikk. Hun valgte elever med lave, gjennomsnittlige og høye

matematiske evner basert på en standardisert aritmetikkprøve. Deretter ble elevenes mestringsforventninger kartlagt, og kategorisert som elever med høy mestringsforventninger og elever med lave mestringsforventninger. Videre ble elevene bedt om å løse vanskelige regneoppgaver. Resultatene viste at elever med høye mestringsforventninger var raskere til å avvise feil løsningsstrategi og løste flere problemer sammenlignet med sine medelever som hadde lave mestringsforventninger, uavhengig om elevene var i gruppa med høyt-presterende elever eller ikke. Høye mestringsforventninger forutså også positive holder til matematikk, noe evner ikke gjorde. I tillegg viste de til Bouffard-Bouchard (1990) som gjorde liknende funn i en studie hvor man undersøkte innflytelsen mestringsforventninger hadde på kognitive prestasjoner, der respondentene hadde lik kunnskap og erfaring innfor et gitt tema. Resultatene viste at studenter som rapporterte om høye mestringsforventninger også hadde høyere ambisjoner, viste større fleksibilitet for valg av strategier og oppnådde bedre resultat, enn studentene som hadde lave mestringsforventninger. Samlet viser disse studiene at elever kan prestere dårlig enten fordi de mangler ferdighetene til å lykkes eller om de har ferdigheten, men mangler mestringsforventningene til å bruke dem optimalt.

2.2.3 Mestringsforventninger og kjønn

Studier som undersøker hvilken rolle kjønn har i elevens vurdering av egne evner, er i følge Chen (2003) motstridene. Analyser fra PISA 2012 viser at det er betydelig kjønnsforskjeller når det kommer til mestringsforventninger i matematikk i Norge, i den forstand at gutter uttrykker sterkere mestringsforventninger enn jenter (Jensen & Nordtvedt, 2013). Dette til tross for at det ikke ble funnet signifikante kjønnsforskjeller i prestasjonsnivået. Til tross for at det ikke var kjønnsforskjeller i resultatene mellom de norske elevene på matematikkprøver i PISA 2012, var det er klart mønster når det kom til holdninger i matematikk (Nordtvedt, 2013). Guttene rapporterte om større grad av motivasjon i matematikk, de rapporterte også om større grad av utholdenhet, kapasitet til problemløsning og høyere selvoppfatning. Det eneste konstruktet innenfor matematikk som jentene skåret høyest på var innenfor matematikkangst. Sammenfallende resultater rapporteres i tidligere studier innenfor mestringsforventninger og kjønn, som også indikerte at gutter har mer tro på egne matematiske ferdigheter enn jenter (Pajares, 2005). I tillegg viste det seg også at gutter hadde mer positive holdning til matematikkfaget. Randhawa et al. (1993) uttrykker lignende resultater fra Canada, i sin studie av 225 elever på high school, som viste at gutter uttrykker høyere forventning om mestring sammenlignet med jenter innenfor matematikk. Resultater fra Pajares og Kranzler (1995) viste derimot ingen eller liten kjønnsforskjell i prestasjon i

matematikk, og heller ikke i mestringsforventninger i matematikk. Heller ikke Chen (2003) fant noen sammenheng mellom kjønn og mestringsforventninger i sine studier, heller ikke i elevenes kalibrering av mestringsforventninger. Pajares og Graham (1999) påpeker imidlertid at jenter og gutter utviser omtrent lik tro på egne evner i matematikk på barneskolen, men allerede innen ungdomsskolen rapporterer guttene sterkere tro på egne matematiske evner sammenlignet med jentene. Samlet sett ser vi at bildet er noe motstridende, men i den retningen at gutter har høyere mestringsforventninger eller at det ikke er noe kjønnsforskjeller. Det er ikke funnet forskningsresultater som viser at jenter har signifikant høyere mestringsforventninger i matematikk enn gutter.

2.3 RAE - Relativ alderseffekt

I mange land, som i Norge, er skolestart sterkt knyttet til fødselsdato. Overgangen til et nytt år fungerer som et skarpt alderskriterium. Dette gir et vilkårlig, men likevel tydelig skille hvor barn som er født like etter hverandre, men på hver sin side av skillelinjen, går på ulike klassetrinn. Alle barn som er født i samme kalenderår starter på skolen samtidig, i august det året de fyller seks år. Med andre ord kan aldersforskjellen mellom elever i samme klasse være opp til tolv måneder.

I Norge er alder og klassetrinn nesten helt sammenfallende da forsering eller repetisjon av klassetrinn nesten ikke eksisterer (Gabrielsen & Lundetræ, 2017). Akselerert eller forsinket skolegang er også lite praktisert. I snitt er det omtrent 400 barn som har utsatt skolestarten til det året de fyller sju, noe som utgjør rundt 0,6% av gjeldene årskull. Så til tross for at «grade retention» er vanlig i flere land, så praktiseres det sjeldent i Norge (Strøm, 2004). Det finnes ingen offisiell oversikt over hvor mange barn som har fremskyndet skolestart, men ifølge Gabrielsen og Lundetræ (2017) er andelen på omtrent 0,2%. I tillegg er andelen av elever som går på private skoler svært liten (Statistisk sentralbyrå, 2020). Dette fører til at i Norge, vil omtrent samtlige barn født innenfor samme kalenderår begynne i grunnskolen samtidig. Det betyr i tillegg at de fleste elever på samme trinn også har like lang skolegang, både i år og måneder. Disse faktorene gjør at data hentet fra norske klasserom er godt egnet til studier som undersøker RAE.

Eksisterende litteratur viser at elever født tidlig i det akademiske året presterer generelt bedre, både innenfor idrett og utdanning, enn elever som er født sent (Aune, Ingvaldsen, Vestheim,

Bjerkeset & Dalen, 2018; Crawford et al., 2013b; Frøseth, Hovdhaugen, Høst & Vibe, 2008; Gabrielsen & Lundetræ, 2017; Gledhill et al., 2002; Lien, Tambs, Oppedal, Heyerdahl & Bjertness, 2005; Olsen & Bjørnsson, 2018; Solli, 2017; Strøm, 2004). Som nevnt i innledningen er RAE konsekvensene av relativ aldersforskjell (Wattie et al., 2008), der relativ alder er en persons alder, sammenlignet med en annen persons alder i den relevante gruppen. Det finnes en rekke teorier og hypoteser om hvorfor relativ alderseffekt eksisterer, men ingen enkel årsak eller kombinasjon av årsaker er til dags dato blitt endelig etablert (Campbell, 2013). Videre påpekes det imidlertid at etter hvert som forskningen innenfor RAE fortsetter, vil det bygges en mer detaljert forståelse om hva RAE er, og hvordan man kan forebygge det. Teorier om årsakene til RAE spenner seg over biologiske og samfunnsvitenskapelige forskningsområder. I det følgende vil jeg kort ta for meg to teorier som omhandler mulige årsaker til RAE, nemlig *biologibasert teori* og *psykososial teori*.

2.3.1 Biologibasert teori

Kanskje den enkleste hypotesen innenfor biologiske årsaker til RAE er *modningshypotesen* (Martin, Foels, Clanton & Moon, 2004). Denne hypotesen handler om at de yngste elevene i alle klasser i gjennomsnitt er ni til tolv måneder mindre nevrologisk modne enn sine medelever. Modningshypotesen hevder at forskjellene i ferdighetsnivå nettopp er et resultat av elevenes variasjon i nevrologisk modning, og at dette kan være en fordel for de eldre elevene og en ulempe for de yngre. Nevrologisk modning gir seg til kjenne i selvregulering, og påvirker funksjoner som oppmerksomhet, metakognisjon og kontroll. Det er kjent at disse funksjonene er relatert til nevrologisk modning og blir mer effektive med alderen. I tillegg, forutsatt lignende vekstrater og pubertetsutbrudd, vil elever i en klasse alltid ha visse fysiske og kognitive forskjeller som varierer innen enhver tolv måneders periode. Et annet område innenfor biologiske faktorer er *prenatale variasjoner*. Det området presenteres kort nedenfor.

På biologisk nivå har det blitt antydnet at prenatale variasjoner kan påvirke fosterutviklingen i livmoren, og påfølgende fremgang etter fødselen (Campbell, 2013). Disse hypotesene baserer seg på forestillinger om at barn som er født på slutten av våren og sommeren har større sannsynlighet for å forstyrrelser i den nevrologiske utviklingen (Martin et al., 2004). Grunnen til dette er at mødrene er i midten av svangerskapet i en periode med større risiko for naturlig fosterutvikling, nemlig i vintermånedene. I følge Martin et al. (2004) tar denne teorien i hovedsak for seg to spesifikke hypoteser, basert på ulik risiko knyttet til svangerskap i vintermånedene. Den første hypotesen omtales som *hypotesen om svangerskapsinfeksjon*, og

omhandler risikoen for infeksjoner, som lungebetennelse og influensa. Dette er infeksjoner som opptrer hyppigere i vintermånedene, og som antas å være assosiert med en atypisk utvikling i fosterets utvikling av sentralnervesystemet. Den andre hypotesen blir referert til som *vitamin D-hypotesen*, og handler i stor grad om mangel på vitamin D under svangerskapet. Dette er en risiko som oppstår i vintermånedene, hvor mor ikke får nok vitamin D på grunn av mangel på dagslys. Det er i henhold til Martin et al. (2004) stadig flere og bedre bevis på at vitamin D spiller en viktig rolle i fosterutviklingen, og at mangel på dette fører til forstyrrelser i fosterutviklingen.

Imidlertid utelukker internasjonale studier fra land med skoleoppstart på andre tidspunkt enn 1. september (England), hvor de yngste elevene fortsatt presterer dårligere enn de eldre elevene, at disse biologiske hypotesene ene og alene ikke kan forklare den relative alderseffekten man ser i opplæringen (Gledhill et al., 2002). I likhet med Campbell (2013), settes disse hypotesene til side da de ikke kan forklare hvorfor man opplever RAE også i det norske skolesystemet. Videre fokuserer jeg på psykososiale forklaringer.

2.3.2 Psykososial teori

På barnets nivå og skolemiljø har psykososial forskning foreslått at umodenhet ved skolestart kan være grunnlaget for variasjon i alderseffekten, ved at relativt yngre og mer umodne elever har en sosial, emosjonell og kognitiv ulempe ved skolestart (Campbell, 2013). Dette er en teori som bygger på samme grunnprinsipper som modningshypotesen, om at eldre elever er mer modne. Martin et al. (2004) fremhever at nevrologisk modning kan ha påvirkning på selvværd og selvfølelse. Hvis det er slik at de eldste elevene både er sterkere og mer modne, påpeker Crawford et al. (2013b) at dette kan ha positiv effekt på både selvtillit, ambisjoner og utvikling av sosiale ferdigheter. Bevis for at yngre elever uforholdsmessig kan få diagnoser som inneholder spesielle læringsvansker på grunnlag av nevrologisk modning, snarere enn basert på en psykologisk eller kognitiv utfordring, støtter denne teorien som en mulig årsak til relativ alderseffekt. Det er ifølge Campbell (2013) en mengde pedagogisk og sosialpsykologisk forskning som støtter denne forestillingen om at merking av et individ som en viss «type», vil kunne påvirke deres reelle oppfatning og atferd. Dette antyder at om en elev blir informert, eksplisitt eller implisitt, om hva denne eleven er stand til å gjøre i et akademisk perspektiv, vil dette ha innvirkning på følgende oppførsel. Videre påpekes det også at elevene ofte godtar rollene som de tildeles, og at dette kan medvirke til at eleven endrer seg i den retningen som forventes ut fra rollen. Forskningen viser også at relativt yngre elever

oftere blir ansett som elever med læringsvansker sammenlignet med de eldre elevene, når de faktisk bare er yngre og i mindre stand til å tilpasse seg de kravene som skolestart krever (Campbell, 2013). Dette påpekes også av Sharp, Hutchison og Whetton (1994) som skriver at barn som er født tidlig på året ofte har gått lengre i førskole-institusjoner, som i barnehagen. Barna som er født sent har derfor mindre førskoleerfaring, og kan dermed være mindre forberedt på de krevende forventningene man møter på skolen. Imidlertid er det flere indikasjoner på at det også er andre faktorer som spiller inn i de yngste elevenes skoleprestasjoner. Disse presenteres nedenfor.

I en undersøkelse gjort av Gledhill et al. (2002) ble det konkludert at lese- og staveferdighetene til de yngste elevene var på det nivået man kunne forvente på den alderen de var. I tillegg fant de at det ikke var forskjell mellom elever født på høsten, våren, vinteren eller sommeren når det kom til gjennomsnittlig IQ eller ved måling av læringsvansker. Til tross for dette hadde de yngste elevene større sannsynlighet for å bli beskrevet som elever med læringsvansker av lærerne sine. Det var også 5,4% større sjanse for at elever født sent i det akademiske året mottok en mild form for spesialundervisning sammenlignet med elevene som er født tidlig (Crawford et al., 2013b). En grunn til dette kan være at de yngre elevene blir sammenlignet med kognitivt eldre og mer modne elever (NOU 2019: 3, 2019). En annen faktor er at det betydelig flere av de yngre elevene som blir diagnostisert med ADHD. Elever som blir oppfattet som elever med læringsvansker har økt risiko for sekundære problem som lav selvtillit, utvikling av emosjonelle problemer eller atferdsproblemer, og har mindre sjanse for å oppnå sitt akademiske potensial (Gledhill et al., 2002). Disse elevene er også mer utsatt for å ha dårligere utviklet sosiale ferdigheter, mindre tro på egne evner og signifikant mindre tro på at deres handlinger kan gjøre noe forskjell.

Baker et al. (2015) fant at lærere oppfattet de eldste elevene som mer faglig dyktig enn sine yngre medelever. De yngste elevene ble også i større grad henvist til mer akademisk støtte, selv om de er på det utviklingsnivået man kunne forvente ut fra alder. Et mer oppsiktsvekkende funn var at elever med atferdsproblem ble oppfattet som akademisk svakere, og at læreren forventet mindre av disse elevene. Læreren forventninger til elevene kan spille en betydelig rolle for hvordan man behandler elevene sine. I en undersøkelse gjort av Jussim et al. (2009) viste det seg at lærere gir mer emosjonell støtte, smiler mer og er mer oppmuntrende mot elever som er forventet å prestere godt. De bruker mer tid med, og gir mer oppmerksomhet til, samt gir flere hint og støtte når disse elevene støter på utfordringer.

Lærerne fokuserte også mer på elevenes prestasjoner ved tilbakemeldinger til de høyt-presterende elevene, mens til de elevene som de forventet mindre av hadde mer fokus på samarbeidsevner og atferd. Det ble også påpekt av Solli (2017) at elevenes sosiale ferdigheter kan være en faktor som beskytter mot lave forventninger om akademisk prestasjon. Hvis det er slik at de eldste elevene er mer modne og det har påvirkning på deres sosiale ferdigheter kan også dette spille en rolle for den relative alderseffekten. Hvis det å være eldre og mer moden har en positiv effekt på læring, er det klart at de eldste elevene har en fordel ved skolestart (Solli, 2017). Dette er også et aspekt når elever blir evaluert på samme tidspunkt, at de eldste barna presterer bedre kun fordi de er eldre og mer modne på prøvetidspunktet (Black, Devereux & Salvanes, 2011; Crawford et al., 2013b).

I følgende tre delkapittel ønsker jeg å se nærmere på forskning på RAE i sammenheng med prestasjoner i matematikk og i sammenheng med kjønn akkurat slik jeg gjorde under kapittel 2.2 om mestringsforventninger. Jeg avslutter kapittel 2.3 med å se på forskning av langtidseffektene av relativ alderseffekt.

2.3.3 Relativ alderseffekt og prestasjon i matematikk

Flere undersøkelser fastslår at RAE er et merkbart fenomen. Gabrielsen og Lundetræ (2017) undersøkte sammenhengen mellom fødselsmåned og leseferdigheter ved hjelp av PIRLS-resultatene fra 2001-2016 på 4.og 5.trinn. Resultatet viste en signifikant forskjell i snittskårene elevene fikk, hvor de eldste i årskullene kom klart best ut. Lignende tendenser er funnet hos 15-16 åringene i Norge, hvor resultatet fra PISA 2000 viste at de yngste elevene på trinnet hadde en betydelig ulempe innenfor lesning sammenlignet med de eldre elevene (Strøm, 2004). I en kartlegging av fire grunnskolekull med til sammen 232 000 elever var det også signifikante forskjeller i gjennomsnittlig grunnskolepoeng mellom de yngste og eldste elevene i kullene (Frøseth et al., 2008). Her ble det påvist en alderseffekt ved utgangen av grunnskolen, slik at de som var født tidlig på året hadde gjennomsnittlig grunnskolepoeng som var 0,15% høyere enn de som var født sent på året. Dette ble forklart med at de eldste elevene fortsatt hadde noe av det forspranget de hadde på de yngste elevene da de startet i grunnskolen, ved at de var eldre og hadde mer erfaring. Lignende resultater er funnet blant elever på 10.trinn i Oslo, hvor den yngste tredjedelen av elevene hadde betydelig dårligere gjennomsnittlig grunnskolepoeng enn tredjedelen av de eldste elevene (Lien et al., 2005). Det ble derimot ikke funnet noen signifikant forskjell mellom elevene i den midterste tredjedelen

og den yngste delen.

Olsen og Bjørnsson (2018) skriver at RAE kan variere på tvers av fagområder, aldersgrupper, studier og land. Som vi ser over er RAE et merkbart fenomen innenfor leseferdigheter, både på barneskolen og i ungdomsskolen, og når det kommer til gjennomsnittlig grunnskolepoeng ved utgangen av grunnskolen. I 2015 gikk Norge fra å gjennomføre TIMSS-undersøkelsen på elever på 4. og 8.trinn til å teste 5. og 9.trinn, noe som resulterte i et dobbelt datasett dette året (Bergem, Kaarstein & Nilsen, 2016). De to påfølgende trinnene besvarte de samme oppgavene, noe som gjorde det mulig å sammenligne resultatene mellom 4. og 5.trinn, og 8. og 9.trinn. Bjerke et al. (til fagfellevurdering) undersøkte sammenhengen mellom fødselsmåned og prestasjon i matematikk ved å ta i bruk dataene fra TIMSS 2015. De fant at det var en signifikant sammenheng mellom RAE og prestasjon innenfor alle emneområdene i TIMSS på 4. 5. og 8.trinn, men ikke for elevene på 9.trinn. De fant også at forskjellen mellom matematikkskår mellom tidlig og sent fødte i samme skoleklasse er signifikant større på 4.trinn enn på 9.trinn. Dette indikerer en nedgang i RAE jo eldre elevene blir (Bjerke et al., til fagfellevurdering). En lignende undersøkelse ble gjennomført hvor fødselsmåned og skår på nasjonale prøver i regning ble sammenlignet. Det ble funnet at alderseffekten var konsistent på 5. 8. og 9.trinn, for både jenter og gutter (Aune et al., 2018). I en analyse av norske resultater fra TIMSS og PISA gjennom 20 år viser det seg at RAE er to ganger større for elever på 4. og 5.trinn enn for elever på 8. og 9.trinn (Olsen & Bjørnsson, 2018). Likevel finner de at estimatene for RAE er substansielt meningsfulle og statistisk signifikante også hos de eldre elevene. Det påpekes imidlertid at for PISA-undersøkelsene for 10.trinn er bilde noe mindre tydelig. Av de 18 analysene på sammenhengen mellom alder ved prøvetidspunktet og skår i matematikk, som ble gjennomført for dataene fra PISA, var kun 11 statistisk signifikante. Likevel understrekes det at tendensen er tydelig. Jo eldre elevene er på prøvetidspunktet, jo bedre presterer de på testene. Denne tendensen påpekes også av Black et al. (2011) som sammenlignet prøveresultater fra skolen etter fødselsmåned med avlagte prøver på sesjon, hvor innkallingen følger fødselsdato, og finner at absolutt alder ved prøvetidspunkt er viktigere enn relativ alder i klasserommet. Med dette antyder de at modning ved prøvetidspunktet kan ha større betydning for prestasjon enn relativ alder i klasserommet.

2.3.4 Relativ alderseffekt og kjønn

Som nevnt over har tendensen vært at gutter presterer bedre på standardiserte tester, mens jenter oppnår bedre eksamenskarakter og standpunkt karakter i matematikk. Studier viser også

at guttene uttrykker større tro på egne evner i matematikk, og har høyere mestringsforventninger enn jentene. Dette til tross for at det ikke ble rapportert kjønnsforskjeller i matematikkprestasjon i hverken PISA 2012 eller TIMSS 2015 (Bergem, 2016; Jensen & Nordtvedt, 2013). At fødselsmåned spiller en vesentlig rolle for skoleprestasjoner er tydelig, og mange av studiene som omhandler RAE, ser også på kjønnsforskjeller knyttet til alder. Cools, Schöne og Strøm (2017) brukte alder ved fullført grunnskole til å estimere alder ved skolestart. Analysene fra denne undersøkelsen viste at rundt tre prosent av barn født i det siste kvartalet fikk utsatt skolestart, og tilsvarende andel av barn født i det første kvartalet fikk fremskutt skolestart. De fant også en kjønnsforskjell i datamaterialet. Av barna født i siste kvartal hadde gutter 2,3% høyere sannsynlighet for å utsette skolestart med ett år enn jenter. På motsatt side, av barna født i 1. kvartal, har guttene 2,2% lavere sannsynlighet for fremskutt skolestart. Det konkluderes også med at utsatt skolestart korrelerer negativt med antall grunnskolepoeng og fullføring av videregående skole. Lignende funn finner man i Solli (2017) som fant at fremskyndet skolestart hadde positiv effekt på å fullføre videregående skole på normert tid. Man ser i tillegg at gutter født i desember har betydelig lavere sannsynlighet for å ha fullført videregående skole innen det året de fyller 19 år (Solli, 2017). Med dette konkluderes det med at ulempen med å være yngst i en klasse er betydelig større for gutter enn for jenter.

Undersøkelser av norske PIRLS-resultater viste en forskjell i leseferdigheter i jenters favør, sammenlignet med jevnaldrende gutter (Gabrielsen & Lundetræ, 2017). Resultatene fra 2001 trekkes særlig frem, hvor jenter født i januar/februar i snitt fikk 55 poeng mer enn gutter født i november/desember. Denne forskjellen tilsvarte mer enn et skoleår. Også Vibe (2008) registrerte en tydelig kjønnsforskjell i grunnskolekarakterer, i jenters favør. I PISA-undersøkelsen for 10.trinn er kjønnsforskjellene moderate, men gjennomgående i den retning at alderseffekten er høyere for gutter (Olsen & Bjørnsson, 2018). Det var imidlertid gjennomgående små kjønnsforskjeller, uavhengig av alder, i alle de nordiske landene i resultatene fra PISA 2012 (Nordtvedt, 2013). Aune et al. (2018) fant derimot at gutter født tidlig i det akademiske året var overrepresentert blant de elevene som presterte høyt i nasjonale prøver i regning, og at jenter født sent var overrepresentert blant de lavt-presterende elevene.

2.3.5 Langtidseffektene av relativ alderseffekt

Crawford et al. (2013b) viser i sin rapport at RAE er størst ved skolestart og synker jo eldre elevene blir. Likevel viser flere undersøkelser at RAE fortsatt eksisterer blant elever i overgangen til videregående opplæring (Frøseth et al., 2008; Strøm, 2004; Vibe, Frøseth, Hovdhaugen & Markussen, 2012). Gabrielsen og Lundetræ (2017) fant noe motstridende resultat i sine undersøkelser av PIRLS-resultatene. Avstanden med hensyn til elevenes snittskår mellom elever født i januar/februar og november/desember var vesentlig større for elevene på 5.trinn enn elevene på 4.trinn. Det påpekes imidlertid at de er forsiktig med å generalisere ut fra dette enkeltresultatet. Til tross for at alderseffekten er størst ved skolestart, er fortsatt gapet mellom de eldste og yngste elevene betydelig ved utgangen av grunnskolen (Frøseth et al., 2008; Lien et al., 2005; Olsen & Bjørnsson, 2018; Strøm, 2004; Vibe et al., 2012). Dette er også en tid hvor elevene står ovenfor valg om videre og høyere utdanning. Elever født sent i det akademiske året har omtrent 2% mindre sjanse for å starte i høyere utdanning ved 18-års alderen, og har 1% mindre sjanse for å fullføre graden sin, enn sine medelever som er født tidlig (Solli, 2017).

Black et al. (2011) konkluderte med at skolestart hadde liten sammenheng med inntekt og utdanningsnivå, og identifiserte heller ingen generell gevinst ved å usette eller fremskynde skolestart for enkeltelever. Likevel utelukkes det ikke at det kan være fordelaktig for den enkelte å gjøre en av delene. Dette støttes også av Sollid (2017) som fant at fremskyndet skolestart hadde positiv effekt på å fullføre videregående på normert tid, men ingen sammenheng med utdanningsnivå eller inntekt. Derimot finner hun sammenheng mellom fødselsmåned og inntekt- og utdanningsnivå. Resultatene viste at barn født tidlig på året hadde større sjanse for å begynne rett i videregående opplæring etter grunnskole og fullføre på normert tid, samt å søke høyere utdanning enn barn født sent på året. De eldste elevene presterer klart bedre enn de yngste, og alderseffekten på snittkarakterene ved utgangen av grunnskolen var nærmest lineær. Karakterene til elevene som var født i desember var i gjennomsnitt 19% lavere enn elever født i januar (Solli, 2017). En sammenheng mellom fødselsmåned og inntekt ved 30 år ble også funnet. Analysene viste at menn født sent på året i snitt tjente 4% mindre enn menn født tidlig i året. Denne sammenhengen ble ikke funnet for kvinner. Til tross for at det finnes store forskjeller når det gjelder utdanning, avhengig av fødselsmåned, finnes det lite bevis for at disse effektene vedvarer i voksen alder (Black et al., 2011; Crawford, Dearden & Greaves, 2013a; Solli, 2017).

Debatten om når elever i Norge skal starte på skolen er fortsatt aktuell. En RAE vil alltid eksistere så lenge skolestart bare skjer en gang i året. Bjerke et al. (til fagfelleevaluering) stiller spørsmål om det finnes en alder for skolestart som gjør ulempen med å være yngst i klassen minst mulig. En elev vil jo alltid være den yngste i sin klasse.

2.4 Forskningsspørsmål og forventende sammenhenger

Tidligere forskning viser at det er sammenheng mellom RAE og prestasjon i skolen. Videre fremmer teori om mestringsforventninger at det er sammenheng mellom mestringsforventninger og prestasjon, og forskningen støtter at det er en positiv relasjon mellom mestringsforventninger i matematikk og prestasjon i faget. Det er likevel ingen dokumentasjon fra Norge som støtter opp om antagelser knyttet til at det er sammenheng mellom RAE og mestringsforventninger i ulike årskull. Denne oppgaven har til hensikt å undersøke om denne korrelasjonen eksisterer. Forskning innenfor RAE viser også at det er signifikante kjønnsforskjeller når det kommer til prestasjon, i den retningen at RAE er sterkest blant gutter, mens forskning på mestringsforventninger viser at gutter rapporterer om høyere mestringsforventninger enn jenter som presterer like godt. Jeg vil derfor undersøke om dette også er tilfelle med mestringsforventninger og RAE, ved å undersøke korrelasjonen mellom RAE, mestringsforventninger og kjønn.

Siden jeg i min litteraturgjennomgang ikke fant noe egnet instrument som måler mestringsforventninger for aktuell aldersgruppe, har jeg i denne masteroppgaven både rapportert om utviklingen og valideringen av et instrument som måler mestringsforventninger blant elever på 9.trinn. Utviklingen ble gjort i samarbeid med de andre deltagerne i forskningsgruppen SESAM. Forskningsspørsmålene og forventede sammenhenger er utarbeidet på grunnlag av teorien og forskningen presentert over. De fungerer også som en utdyping og konkretisering av problemstillingen jeg presenterte i kapittel 1.4. Dermed er følgende forskningsspørsmål formulert for denne masteroppgaven:

- Hvordan vil et instrument som måler mestringsforventninger i matematikk hos elever på 9.trinn kunne se ut?
- Hvordan er nivået på mestringsforventninger i matematikk sett i forhold til relativ alder?
- Er det forskjeller i rapportert mestringsforventning til de ulike items sett i lys av emneområdene fra TIMSS, og relativ alder?

- Er det kjønnsforskjeller i rapportert mestringsforventning i matematikk og RAE. Eventuelt hvilke forskjeller er det?
- Er det forskjell mellom høyt-presterende og lavt-presterende elevers mestringsforventning i matematikk knyttet til relativ alder? Hvilke forskjeller er det?

Det første forskningsspørsmålet vil bli viet ekstra stor plass i denne masteroppgaven siden utvikling og validering av et slik instrument er en omfattende prosess. Denne prosessen blir også godt synlig i metodedelen som følger i neste kapittel.

3. Metode

Forskningsspørsmålene mine la føringer for valg av metode for innhenting av data og for hvordan disse dataene skulle analyseres. Innledningsvis i dette kapitlet vil jeg redegjøre for valgt metodisk tilnærming, med fokus på spørreundersøkelse som forskningsmetode. Gjennomgående i dette kapitlet legges det vekt på hvordan forskningsgruppa SESAM utviklet et instrument som måler mestringsforventninger i matematikk hos elever på 9.trinn. Nyutvikling av et slik måleinstrument krever pilotering og grundige, stegvise analyser. Derfor beskrives utviklingen av dette instrumentet detaljert. For å kunne redegjøre for nødvendige analyser, blir prinsippet bak Rasch-analyse viet et eget delkapittel. Rasch-modellen blir brukt til å validere instrumentet og senere til å analysere dataene fra hovedundersøkelsen. Videre blir hovedundersøkelsen og populasjonen presentert, før det avslutningsvis gjøres rede for de etiske betraktningene knyttet til oppgaven. Gjennomgående i oppgaven vil begrepet spørreundersøkelse bli brukt når det er snakk om hele undersøkelsen. Spørreundersøkelsen består videre av tre måleinstrumenter, hvor hvert måleinstrument har til hensikt å belyse henholdsvis; *mestringsforventninger*, *kilder til mestringsforventninger* og *mestring* i matematikk. Måleinstrument eller instrumentet vil bli brukt når det kun er snakk om en av disse delene av undersøkelsen.

Denne masteroppgaven baserer seg i all hovedsak på instrumentet som har til hensikt å måle mestringsforventninger i matematikk hos elever på 9.trinn, og dermed vil det vies det mest oppmerksomhet til utviklingen og valideringen av dette instrumentet. Forskningsspørsmålet som omhandler høyt- og lavt-presterende elever baserer seg på undersøkelsens siste del, hvor elevene løste oppgavene de hadde fått presentert i instrumentet for mestringsforventninger. For å kategorisere elevene som høyt-presterende eller lavt-presterende ble skår på mestring telt opp, og elevene som fikk 1-2 oppgaver riktig ble kategorisert som lavt-presterende og elevene som fikk 7-12 oppgaver riktig ble kategorisert som høyt-presterende. Instrumentet som måler mestringsforventninger er gjennomgått en valideringsprosess, med en pilot-studie (N=41) og hovedstudien (N=583). Analyser og resultat av pilotstudien la føringen for hvordan den endelige utgaven av mestringsforventningsinstrumentet ble. Analyser knyttet til pilotstudien og utviklingen av instrumentet presenteres i dette kapitlet, og valideringen av det endelige instrumentet presenteres i resultatkapitlets første del.

3.1 Metodisk tilnærming

Samfunnsvitenskapelig forskning skiller i hovedsak mellom to forskningsmetoder; kvalitativ og kvantitativ metode (Johannessen, Tufte & Christoffersen, 2012). Kvantitative samfunnsvitenskapelige undersøkelser forteller oss ofte noe om utbredelsen av ett eller flere fenomen og sammenhengen mellom fenomener, og baserer seg ofte på et stort utvalg respondenter (Tufte, 2011). En kvantitativ tilnærming er ofte teoristyr, og forskeren stiller spørsmål og avleder hypoteser fra det teoretiske perspektivet som er relevant for de fenomenene som undersøkelsen har til hensikt å studere (Ringdal, 2013). Kvalitativ forskning derimot involverer færre respondenter, og fokuserer mindre på utbredelsen og mer inngående på selve fenomenet. Slike data fremstilles ofte i form av tekster, bilder og lyd, og legger vekt på en fortolkning av dataene (Johannessen et al., 2012).

Vår undersøkelse er det metodelitteraturen kaller en tverrsnittstudie. En tverrsnittstudie baserer seg på en spørreundersøkelse på et stort utvalg i en begrenset tidsperiode, hvor hver respondent bare spørres én gang. En slik studie gir bare et øyeblikksbilde av respondentene, og kan dermed ikke brukes til å gjøre slutninger om prosesser som kan variere over tid (Ringdal, 2013). Cohen, Manion og Morrison (2018) påpeker at et større utvalg som grunnlag for kvantitative undersøkelser gir bedre egnet datagrunnlag for å undersøke utbredelsen av fenomener. Hovedregelen er at undergrupper i et utvalg må være representert av minimum 30 respondenter, der en helst ser at antallet ikke går under 100 respondenter (Johannessen et al., 2012). Får å oppnå minimum 100 respondenter i hver undergruppene, var jeg avhengig av en stor andel respondenter totalt i undersøkelsen. For å nå ut til flest mulig respondenter ble medlemmene i SESAM enige om å ta i bruk en kvantitativ spørreundersøkelse som forskningsmetode.

For å besvare oppgavens forskningsspørsmål var det avgjørende å innhente data fra elever født tidlig og sent i det akademiske året. For å undersøke om det er sammenheng mellom elevs mestringsforventning og relativ alder har jeg, sammen med to medstudenter, samlet kvantitative talldata. Gruppen SESAM utviklet en spørreundersøkelse som besto av tre måleinstrumenter, og vi som medlemmer stod fri til å bruke de dataene fra datainnsamlingen som belyste hver av gruppemedlemmenes respektive forskningsspørsmål. Instrumentet som måler mestringsforventninger ble utviklet av gruppemedlemmene i SESAM da vi ikke fant et eksisterende måleinstrument som egnet seg til bruk på 9.trinn i Norge. Instrumentet som har til hensikt å belyse kildene til mestringsforventningene baserer seg på Usher og Pajares

(2009) sitt instrument, og er oversatt til norsk av medlemmene i SESAM. Det siste instrumentet måler mestring i matematikk. Utviklingsprosessen for disse instrumentene, og spørreundersøkelsen i sin helhet presenteres nedenfor, i kapittel 3.4. På bakgrunn av at vi var tre studenter som deltok i samme forskningsprosjekt, og ønsket å samle inn data sammen var vi avhengig av å utvikle en spørreundersøkelse som tok utgangspunkt i tre ulike problemstillinger. Siden undersøkelsen denne oppgaven baserer seg på også er utgangspunktet for to andre masteroppgaver, vil spørreundersøkelsen bestå av instrumenter som ikke er relevant for min masteroppgave, og som det derfor vil vies mindre plass til.

3.1.1 Spørreundersøkelse som forskningsmetode

En spørreundersøkelse er en systematisk måte å samle inn data fra et utvalg respondenter. I følge Johannessen et al. (2012) er en vanlig oppfatning at det er relativt enkelt å utforme et spørreskjema, og et arbeid det ofte tas for lett på (Kleven & Hjordemaal, 2018). Men å utvikle et spørreskjema er både tidkrevende og hardt arbeid. Utgangspunktet for utvikling av spørreskjema er undersøkelsens problemstilling, og forskningsspørsmål. Det er derfor helt vesentlig at spørsmålene som utgjør spørreundersøkelsen belyser problemstillingen. Det finnes ingen fasit på hvordan et spørreskjema skal utformes, men det er noen sider ved utviklingsprosessen en forsker må være bevisst på. Disse presenteres nedenfor.

Ved gjennomføring av en spørreundersøkelse er det viktig å unngå for mye belastning på respondentene. Spørreskjemaet må ha tydelige og entydige spørsmål som er enkle å forstå, og det bør ikke ta for lang tid å fullføre undersøkelsen (Cohen et al., 2018). For mye belastning kan føre til dårlig kvalitet eller feilaktig respons, manglende respons eller at respondentene gir opp. Cohen et al. (2018) bemerker også at å fylle ut et spørreskjema kan være mentalt krevende, og forskeren må være bevisst hvilken innsats det krever av respondentene for å fullføre undersøkelsen. Dette var aspekter vi i gruppen SESAM diskuterte nøye da vi utviklet måleinstrumentene som utgjorde spørreundersøkelsen.

Det er mange typer spørreundersøkelser, men en enkel tommelfingerregel er at jo større utvalget er, jo mer strukturert bør undersøkelsen være (Cohen et al., 2018). Vi i SESAM hadde som mål å innhente data fra et stort utvalgt, og dermed besto spørreundersøkelsen vår i all hovedsak av lukkede spørsmål. Lukkede spørsmål forskriver de svarene som respondentene kan velge fra. Slike spørsmål er ofte mer direkte og bevisst mer fokuserte enn åpne spørsmål, noe som gjør at det kan oppleves enklere for respondentene å besvare

(Ringdal, 2013). Strukturerte, lukkede spørsmål er også nyttige fordi de kan generere responsfrekvenser som er mottakelige for statistisk behandling og analyse. De gjør det også mulig for sammenligning mellom undergrupper i utvalget, som var utgangspunktet i min problemstilling. I tillegg er de raskere å kode og analysere enn tekstrike data. På en annen side har ikke respondentene mulighet til å velge svar utenfor de standardiserte svarkategoriene og man kan da miste viktig informasjon, noe som kan være en svakhet ved dataene (Kleven & Hjordemaal, 2018).

Elevens mestring, mestringsforventning, og kildene til disse, er latente variabler som ikke kan observeres direkte, men måles ved hjelp av konkrete uttrykk. Ved måling av mestringsforventninger i matematikk må elevene vurdere hvor trygge de er på å mestre den spesifikke matematikkoppgaven, og må derfor ha mulighet til å gradere svarene sine. I likhet med Usher og Pajares (2009) valgte vi å ta i bruk en seksstegs Likert-skala som svaralternativ. En Likert-skala er en skala med flere verdier, hvor respondentene vurderer grad av enighet til ulike utsagn. Slike skalaer er mye brukt i forskning, og med rette, for de kombinerer muligheten for en fleksibel respons med evnen til å fastslå frekvenser, korrelasjon og andre former for kvantitative analyser (Cohen et al., 2018). Mestringsforventninger angår personers oppfatning av egne evner til å oppnå et gitt resultat, og måles ved grad av sikkerhet. For at respondentene skal få mulighet til å nyansere svaret sitt, og dermed gjenspeile deres reelle oppfatning var det nødvendig å ta i bruk en skala som ga respondentene er mulighet til en slik vurdering. Det finnes ikke et absolutt svar på hvor mange verdier en Likert-skala skal bestå av, men å ha minst fem verdier gir oss mulighet til å gjøre mer avanserte og omfattende analyser enn det er mulig med få verdier (Johannessen et al., 2012). Vi valgte seks svarkategorier i vår undersøkelse. Bakgrunnen for dette valget baserte seg på at vi ønsker like svarskalaer på de ulike instrumentene, og instrumentet utviklet av Usher og Pajares (2009) tok i bruk seks svarkategorier. Denne seksstegs-skalaen ble derfor brukt i instrumentet for å måle mestringsforventninger og kilder til mestringsforventninger, altså undersøkelsens første og andre del. Bandura (2006) påpeker, i likhet med Cohen et al. (2018), at respondenter ofte ikke velger de «ekstreme» alternativene. I tilfeller med skalaer som tar i bruk fire svaralternativer, vil respondentene i ytterste konsekvens kun ha to alternativer å velge mellom. Dette var også et argument for å ta i bruk en svarskala med seks alternativer.

Selv om slike skalaer er mye brukt i forskning har de også sine begrensninger. Det er ikke mulig å anta at intervallene mellom de ulike kategoriene skalaen består av er like (Cohen et

al., 2018). I beskrivelsen av svarskalaen i selve spørreundersøkelsen ble det kun påpekte hva alternativ én og alternativ seks stod for (se Vedlegg 1). Dermed ble det opp til respondentene selv å avgjøre graden av trygghet og sikkerhet på de andre alternativene. Dette fører til at man ikke med sikkerhet kan si at de tolker alternativene likt. Derav kan man ikke fastslå at en vurdering av verdi fire er dobbelt så kraftfullt som verdi to. Med andre ord kan man ikke ta for gitt at følelsesintensiteten mellom *veldig enig* og *enig* på en eller annen måte samsvarer med *veldig uenig* og *uenig*. Dette er ifølge Cohen et al. (2018) ulovlige slutninger. Det er også viktig å være klar over at tallverdier kan ha ulik mening for forskjellige mennesker. En person kan bruke et bestemt kriterium for å tildele en skår fem på en sekspunktsskala, mens en annen person kan bruke nøyaktig samme kriterier vil tildele en skår på fire på akkurat samme skala.

Slike skaleringsproblemer som nevnes ovenfor behandles mer presist i moderne testteori enn i klassisk testteori. Rasch-modellen gir oss et rammeverk som eksplisitt undersøker denne antagelsen, og parametiserer intervallene som definerer kategoriene uten antagelse om at kategoriene er av samme størrelse (Engelhard Jr, 2013). Rasch Rating Scale Model (RSM) er velegnet for polytome Likert-skalaer, og er den modellen som vi bruker i denne masteroppgaven. Dette er en av grunnene til at moderne testteori, som Rasch-modellen tilhører, stadig brukes mer i utvikling og validering av psykologiske tester (Friborg, 2010). Før jeg vil gjøre rede for begrepene validitet og reliabilitet, og før utviklingen av instrumentene forklares mer i detalj, er det nødvendig å få et dypere innblikk i hva Rasch-modellen og Rasch-analyse er, og hvorfor denne testteorien var hensiktsmessig å ta i bruk i denne masteroppgaven. Dette presenteres nedenfor.

3.2 Rasch Rating Scale Model

Moderne testteori kom på 1960-tallet for å bøte på noen av problemene med klassisk testteori (Friborg, 2010). Rasch-analyse er en moderne testteori som ble utviklet for å forbedre presisjonen som forskere konstruerer instrumenter med, kontrollere kvaliteten til et instrument og beregne respondenters prestasjon (Boone, 2016). Rasch-teknikker har blitt mer vanlig, men mange forskere bruker fortsatt klassisk testteori til å utvikle og validere instrumenter.

Valideringen av måleinstrumentet for mestringsforventninger og ytterligere analyser knyttet til hvert av de ulike forskningsspørsmålene gjøres ved hjelp av RSM. Det engelske begrepet *item* vil gjennomgående bli brukt i masteroppgaven, da jeg ikke finner en fullverdig oversettelse til norsk. *Items/items* kan være en oppgave, et spørsmål eller en påstand. På

samme måte vil de engelske begrepene infit, outfit og misfit vil bli brukt for å diskutere kvaliteten til målingene som er gjennomført. Disse begrepene forklares nærmere nedenfor når jeg vil se nærmere på Rasch-modellen, og måleenheten logit.

3.2.1 Rasch-modellen og logits

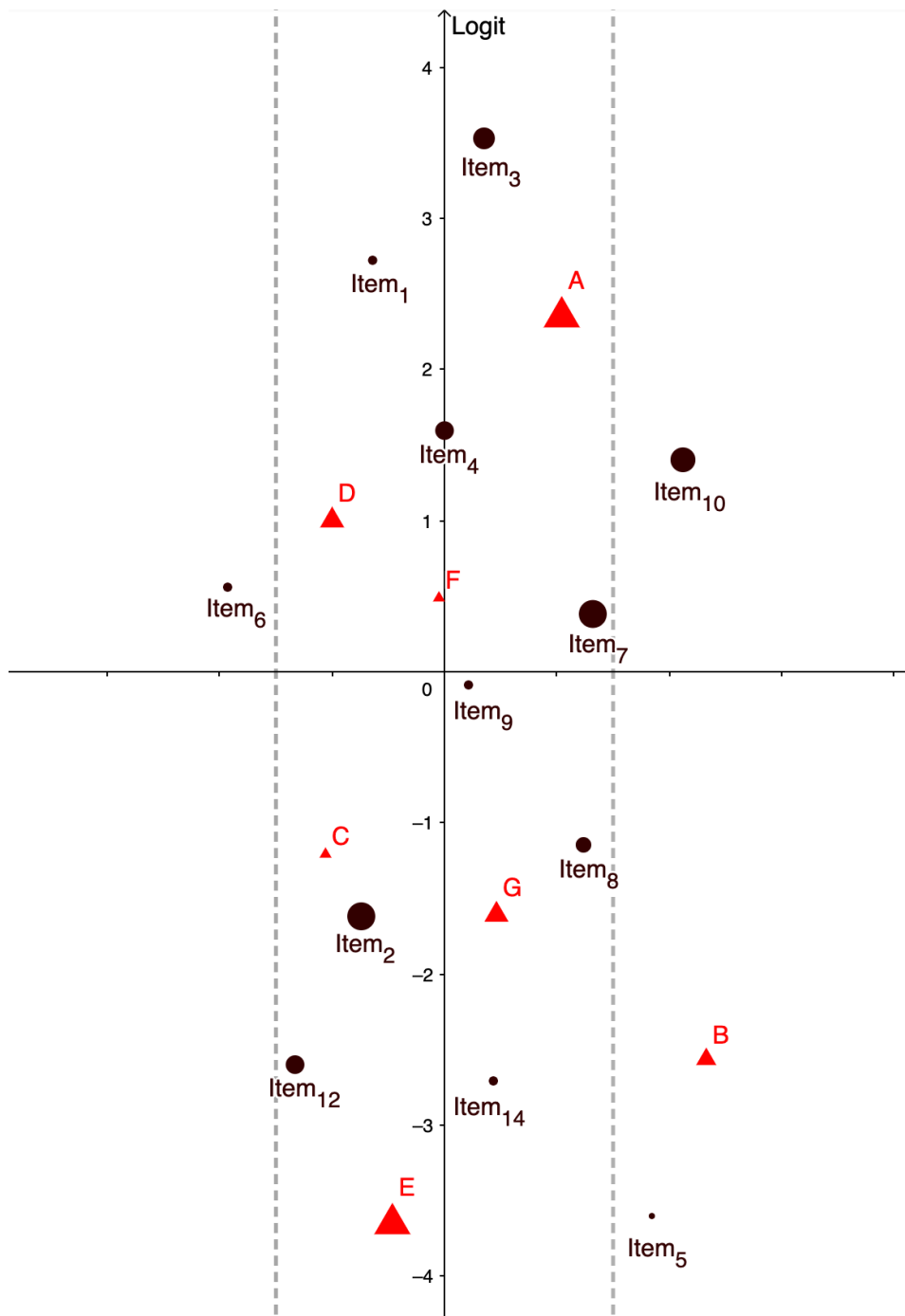
Rasch-modellen er en teoretisk matematisk beskrivelse av hvordan grunnleggende målinger kan fungere med sosiale, latente variabler (Bond & Fox, 2015). Rasch-analyse brukes ofte for å måle egenskaper som ikke trer klart frem i dagen og som kan virke i det skjulte, som evner eller holdninger (Glen, 2016). For å gjøre dette brukes en Rasch-modell, som på mange måter kan oppleves enklere og mer intuitiv enn andre modeller. En Rasch-modell minner om en linjal, men istedenfor centimeter måler skalaen items ved hjelp av såkalte *logits*. Innsikt i de underliggende enhetene som brukes i logit-skalaen er viktig for å kunne forstå Rasch-modellen (Engelhard Jr, 2013). Logit er en ikke-lineær transformasjon av proporsjoner som brukes til å lage en lineær-skala. Dette gjør at den fungerer bra å bruke på Liker-skala data. Med andre ord er logit-skalaen en intervallskala, hvor intervallene mellom hvert punkt på variabelkartet har en konsistent verdi eller betydning. Jo høyere en person er lokalisert i variabelkartet, desto mer kompetent oppfatter de seg selv til å være. Det samme gjelder for items. Ut fra dette Leser vi at nivået på items og personer kan leses av på samme skala. Desto høyere oppfattet vanskelighetsgrad itemet har, desto høyere opp vil det være på variabelkartet. Tilsvarende desto høyere prestasjonsevne på personen som har svart på instrumentet, desto høyere oppe på variabelkartet vil personen finne seg. Et variabelkart gir på denne måten en visuell visning av den underliggende latente variabelen i et format som ligner vår ideelle visjon om et måleinstrument (se Figur 1) (Engelhard Jr, 2013)

Rasch-modellen setter rutinemessig 50% sannsynlighet for suksess for enhver respondent som er lokalisert på samme nivå som oppgaven på logit-skalaen (Bond & Fox, 2015). Med andre ord vil respondenter som befinner seg under et item i variabelkartet ha mindre enn 50% sannsynlighet til å lykkes med den oppgaven. Hver person og hvert item er lokalisert langs logit-skalaen på bakgrunn av en estimert verdi. Ingen reelle empiriske data vil noen gang passe perfekt til Rasch-modellens teoretiske ideal. Spørsmålet vil heller dreier seg om dataene passer godt nok til å støtte de beslutningene vi ønsker å ta (Bond & Fox, 2015).

Estimeringsprosedyren i Rasch-analyser har to forskjellige faser; den første om kalibrering av vanskelighetsgrad og den andre om estimering av fit (Bond & Fox, 2015). I Rasch-analysens

første fase sier Rasch-modellen at det er to parameter som bestemmer hvor godt en person klarer å svare på en oppgave; personens kompetanse og oppgavens vanskelighetsgrad (Granger, 2008). Siden denne oppgaven dreier seg om mestringsforventninger, er det mer logisk å omtale dette som; *elevens opplevelse av egen kompetanse* og *itemets oppfattede vanskelighetsgrad*. Begge disse måles med logit i variabelkartet. Rasch-analyse transformerer dermed ordinale datamateriale, til kontinuerlige data slik at de kan analyseres ved for eksempel t-test.

Rasch-modellen gir også indikasjoner for å hjelpe forskeren å avgjøre om det er nok spredning langs den ideelle linjen, i motsetning til klumper av dem i variabelkartet. Figur 1 viser en illustrasjon av de ideelle linjene i Rasch-modellen, med god spredning i item og personer. I Rasch-analyser kommer dette frem ved hjelp av to målinger, personreliabilitet og itemreliabilitet. Personreliabilitet indikerer sjansen for at estimatene for respondentenes score vil gjenta seg, gitt at de gjennomfører et sett av oppgaver som måler det samme konstruktet. Høy personreliabilitet betyr dermed at man har utviklet en hendelseslinje der noen personer vil skåre høyt og andre vil skåre lavt, og at vi kan forvente konsistens av disse slutningene. Itemreliabiliteten indikerer generaliserbarheten til oppgavens eller påstandens plassering langs linjen hvis disse oppgavene ble gitt til et annet, men lignede utvalg. Høy skår på denne innebærer at noen oppgaver oppfattes som vanskelige og noen oppgaver som lette, og at vi kan forvente at også disse slutningene vil være konsistente.



Figur 1- Illustrasjon av variabelkartet med de ideelle linjene. Illustrasjonen viser god spredning av items og personer.

3.2.2 Invarians og endimensjonalitet

“You can’t measure change with measure that changes” er et sitat som ofte brukes om invariante målinger (Bond & Fox, 2015, s. 83). Invarians er en avgjørende egenskap ved alle vitenskapelige målinger, og innebærer at funksjonen til et måleinstrument forblir uendret på tvers av alle egnede bruksområder. I humanvitenskapen kan vi studere forskjellige variabler, men prinsippet om invarians bør likevel gjelde. Basert på det teoretiske og filosofiske rammeverket til Rasch-modellen er det utviklet fem essensielle krav for at en måleprosess skal være invariant (Engelhard Jr, 2013, s. 43).

1. Måling av personer må være uavhengig av de items som tilfeldigvis brukes til målingen.
2. En mer kompetent person må alltid ha større sjanse for å lykkes på et item enn en mindre kompetent person.
3. Kalibrering av de items som brukes må være uavhengig av de personene som brukes til kalibreringen
4. Enhver person må ha større sjanse for å lykkes på et enkelt item, enn på et vanskelig item
5. Items og personer må være lokalisert samtidig på en enkelt underliggende latent variabel i variabelkartet.

I følge Engelhard Jr (2013) representerer disse fem kravene et sett av regler for å definere idealtypen av målinger. Disse kravene er definert *a priori*, og kommer dermed ikke fra et bestemt datasett. Denne listen presenterer et sett med krav, heller enn antagelser, fordi instrumenter må oppfylle eller tilnærmet oppfylle disse vilkårene for å gi nyttige slutninger som kan generaliseres. Krav én og tre tar for seg kravene for å kunne generalisere utover den bestemte gruppen personer som brukes til å kalibrere items(ene), samt generalisere utover de bestemte items eller observasjoner som brukes til å måle personene. For krav to og fire er invariansen til et måleinstrument knyttet til at vanskelighetsgraden til en oppgave skal oppleves likt av alle respondenter uavhengig av kjønn, alder og sosioøkonomisk bakgrunn (Linacre, 2002). For å måle dette brukes en DTF- analyse (Differential Test Functioning) eller en DIF-analyse (Differential Item Functioning). DIF og DTF har lenge vært anerkjent som en potensiell kilde til bias i måleprosessen. Bias i en undersøkelse innebærer at slutningene eller resultatene i en studie blir påvirket av systematiske skjevheter på tvers av grupper i undersøkelsen (S. Grønmo, 2020).

For å validere måleinstrumentet utviklet for å måle mestringsforventninger i matematikk hos elever på 9.trinn vil en DTF-analyse tas i bruk. DTF-analyser har til hensikt å identifiserer om det finnes signifikante forskjeller, på tvers av grupper, av andelen respondenter som har svart korrekt på et gitt item, og som tilsynelatende har samme evner på (Tennant & Pallant, 2007). Hvis en test eller et instrument måler den samme evnen på samme måte på tvers av grupper, sett bort fra tilfeldige variasjoner, bør suksessraten være uavhengig av de ulike gruppesammensetningene. Med andre ord gjør DTF-analyser det mulig å utforske om instrumentet favoriserer en gruppe i utvalget over en annen. I slike undersøkelser er det viktig å skille mellom reelle forskjeller i oppfattede ferdigheter og bias i undersøkelsen (Utdanningsdirektoratet, 2018). Reelle forskjeller er ofte det mange undersøkelser har til hensikt å måle, men samtidig må man utelukke bias som kan føre til feilaktige slutninger og ugyldige resultater.

Det femte kravet om invarians er direkte knyttet til en annen forutsetning for Rasch-modellen, nemlig endimensjonalitet. Endimensjonalitet er en forutsetning for at en måling skal kunne være invariant, og handler om at et instrument må være utformet for å representere én latent variabel om gangen. En god målingsprosess i humanistiske fag vil gi estimeringer om én egenskap om gangen, og ikke utilsiktet forveksle to eller flere menneskelige egenskaper i ett mål eller poengsum. Endimensjonalitet undersøkes med «item polarity» i fit-statistikken (Bond & Fox, 2015). Denne statistikken forteller om responsen på et gitt item stemmer overens med respondentenes opplevelse av egne evner og itemets oppfattede vanskelighetsgrad. Korrelasjonsverdier som er negative eller nær null signaliserer problematiske items som ikke er konsistent med det konstruert som måles (Bond & Fox, 2015). ICC- Item Characteristics Curves kan også vise om måleinstrumentet inneholder en ekstra dimensjon. En ICC viser Rasch-modellens forutsigelse for hvert mål knyttet til hvert items oppfattede vanskelighetsgrad.

3.3.3 Fit

I Rasch-målinger er «fit» en kvalitetskontroll. Fit-statistikk gir en indikasjon på om forskeren har fullført en oppgave av tilstrekkelig kvalitet for å tillate at verdien for personer og items kan vises med intervallnivå. Rasch-analyser indikerer hvor godt hvert item passer inn i det underliggende konstruert (Bond & Fox, 2015). Items som ikke passer i konstruert, den ideelle rette linjen, er de som avviker fra det som forventes. Det er ikke slik at alle items skal

ligge direkte på denne linjen, men de må være nærme nok. Hvis noen items havner utenfor, enten til høyre eller venstre for linjen, følger de ikke samme mønster som de andre items. I måling av elevers oppfattede evner vil det være mest hensiktsmessig å ekskludere disse, fremfor å omformulere dem (Bond & Fox, 2015).

For å måle fit brukes infit og outfit MNSQ, og infit og outfit ZSTD (Bond & Fox, 2015). ZSTD er en standardisert t-test, og indikerer hvor sannsynlig misfiten eller avviket i datamaterialet er. ZSTD kan være både positive eller negative verdier, hvor negative verdier indikerer mindre variasjon enn modellen forutså og positive verdier indikerer mer variasjon. Den forventede verdien til ZSTD skal være tilnærmet 0, og verdier fra -2 til 2 regnes som akseptable. Rasch-analyseprogrammer rapporterer vanligvis fit-statistikken som to chi-kvadrat-forholdstall; infit MNSQ og outfit MNSQ. Outfit baserer seg på den konvensjonelle summen av de kvadrerte standardiserte residualene. Residual representerer differansen mellom Rasch-modellens teoretiske forventning til hvert items opptreden, og den opptreden som faktisk oppsto i selve datamatriksen. Infit MNSQ er på en annen side en informasjonsvektet sum. Denne verdien er basert på chi-kvadratstatistikken med hver observasjon vektet av sin statistiske informasjon (Linacre, u.d.). Den statistiske informasjonen i en Rasch-observasjon er dens modellvarians, det vil si kvadratet til modellens standardavvik for observasjonen og dens forventede Rasch-verdi. Dette avviket er større for godt tilpassede observasjoner, og mindre for ekstreme observasjoner. Effekten av dette gjør infit mindre påvirket av outliers, og mer sensitiv overfor mønstre av iliggende observasjoner (inlying observations). Bond og Fox (2015) påpeker imidlertid at problemer i datamaterialer som er forårsaket av stor outfit vanligvis er enklere å identifisere, diagnostisere og rette opp enn problemer forårsaket av store infit-verdier. Videre skrives det også at infit er en større trussel mot målingens validitet, enn det outfit er. Linacre (u.d.) anbefaler å ikke rapportere infit med mindre dataene er sterkt forurenset med irrelevant outliers. På bakgrunn av dette vil jeg gjennomgående i denne masteroppgaven diskutere måleinstrumentets kvalitet med hjelp av outfit MNSQ.

Både infit- og outfit MNSQ adresserer spørsmålet «hvor mye misfit eller avvik eksisterer i datamatriksen?». Et eksempel; hvis outfit MNSQ verdien er mindre enn én, si 0,78, indikerer dette 22% mer variasjon i det observerte mønstret enn det Rasch-modellen forutså. For MNSQ forventer man at verdiene skal være nært 1, og i undersøkelser som baserer seg på Likert-skalaer tar man utgangspunkt i at outfit MNSQ-verdien skal ligge innenfor 0,6-1,4

(Bond & Fox, 2015, s. 273). Om MNSQ-verdien til et item er større enn 1,4 betegnes det som «underfit», og vil havne på høyre side av den ideelle linjen. Dette kan senke kvaliteten til påfølgende målinger. Om MNSQ-verdien er mindre enn 0,6 betegnes det som «overfit» og plasseres på venstre side av midtlinjen. Om et item er overfit omtales det som for godt til å være sant, og kan villedes til å konkludere at målingene er bedre enn det de egentlige er. Dette illustreres i Figur 1, hvor to item er underfit og ett item er overfit. I praktiske forskningssituasjoner som inkluderer mennesker, er det lite sannsynlig at dette har noen praktisk betydning i det hele tatt. Som oppsummering vil fit-indikatorer bruke mønsteret av responsen til å estimere hvor mye misfit som eksisterer i datamaterialet (MNSQ) og hvor sannsynlig denne misfiten er (ZSTD).

3.3 Validitet og reliabilitet

Før utvikling av instrumentene beskrives, vil jeg kort gjøre rede for begrepene validitet og reliabilitet. For at forskere og andre yrkesutøvere skal kunne ta holdbare avgjørelser er det viktig at datamaterialet er gyldig og reliabelt. Tre egenskaper som benyttes til å vurdere kvaliteten til målinger er dimensjonalitet, reliabilitet og validitet (Ringdal, 2013). Som tidligere nevnt har jeg brukt Rasch-analyse for å validere instrumentet som måler meststringsforventninger. Dimensjonalitet handler om at målingene baserer seg på et avgrenset teoretisk begrep, og ikke kan splittes i flere underdimensjoner. Dette er et av hovedaspektene i Rasch-modellen, og beskrives nærmere nedenfor. Reliabilitet er knyttet til hvor pålitelige målingene er, og går ut på om gjentatte målinger med samme måleinstrument gir tilstrekkelige samsvarende resultat (Ringdal, 2013). For å belyse om en vitenskapelig problemstilling er pålitelighet, er høy reliabilitet en forutsetning. Reliabilitetsindeksen er standardisert og går fra null til én, hvor null representerer ingen pålitelighet og én er full pålitelighet (Friborg, 2010). Hovedregelen for disse verdiene er at ratioer som er større enn 0,70 blir ansett som akseptable og større enn 0,90 anses som utmerket. I følge Halvorsen (2008) er høy reliabilitet en forutsetning for at dataene skal kunne brukes til å undersøke problemstillingen, men det er ikke en tilstrekkelig forutsetning alene. Dataene må i tillegg være valide, altså gyldige.

Mens reliabilitet er et rent empirisk spørsmål, krever validitet også en teoretisk vurdering (Ringdal, 2013). I motsetning til reliabilitet, som ofte kan uttrykkes med tall, er validitet et gradsspørsmål. Mens man tidligere anså validitet som en egenskap ved måleprosessen, er det

nå vanlig å betrakte validitet som grunnlaget for å kunne tolke dataene som et uttrykk for det begrepet man har operasjonalisert (Kleven & Hjordemaal, 2018). Konstruktvaliditet handler om det er samsvar mellom konstruktet slik det er definert teoretisk, og konstruktet slik vi har operasjonalisert det. I vurdering av konstruktvaliditet blir måleproblemer tatt på alvor, i den forstand at en ikke tar for gitt at målingene er riktig, heller ikke at målingene definerer det teoretiske begrepet som er brukt. Kleven og Hjordemaal (2018) understreker at det aldri vil være fullstendig samsvar mellom det teoretiske konstruktet og slik man har operasjonalisert det, og at validitet derfor i stor grad handler om hvor godt eller dårlig det er rimelig å tro at dette samsvaret er.

Indre validitet brukes utelukkende knyttet til tolkningen av relasjon mellom variablene, slik de har blitt operasjonalisert i problemstillingen (Ringdal, 2013). God indre validitet medfører mulighetene til å stole på slutninger om relasjoner og årsakssammenhenger mellom variabler. Det påpekes likevel at positiv relasjon mellom variabler slett ikke er et bevis på at det foreligger årsakssammenhenger mellom variablene. En årsaksrelasjon oppstår dersom en variabel har en viss påvirkning på den andre variabelen, men innebærer ofte ikke hvorfor denne påvirkningen oppstår. Ytre validitet handler om resultatenes generaliserbarhet. Altså for hvilke personer resultatene er gyldig for og i hvilke situasjoner de er gyldig i (Kleven & Hjordemaal, 2018) For å kunne regne resultatene som overførbare til undersøkelsens populasjon, er det viktig at utvalget er representativt for populasjonen.

Det er ifølge Kleven og Hjordemaal (2018) en vanlig misforståelse at reliabilitet blir prioritert i kvantitativ forskning, mens validitet får mest fokus i kvalitativ forskning. Det vil derfor være viktig å påpeke at også innenfor kvantitativ forskning er konstruktvaliditet et kriterium for gode måleresultater. Reliabilitet er viktig fordi det påvirker konstruktvaliditeten, og lav reliabilitet bidrar til å svekke validiteten. Mens reliabilitet påvirkes av tilfeldige målefeil, vil systematiske målefeil direkte gå utover dataenes validitet (Halvorsen, 2008). Likevel kan ikke måleresultat som er forårsaket av tilfeldige være noen valid måling i det store og hele. Prinsippene om valide og reliable målinger er viktig å ha i tankene ved utvikling av måleinstrumenter og spørreundersøkelser.

3.4 Utvikling av måleinstrumentene

Forskningsprosjektet hadde til hensikt å undersøke elevens mestringsforventninger, kilder til mestringsforventninger og prestasjon i matematikk. Spørreundersøkelsen besto dermed av tre måleinstrumenter og bakgrunnsvariablene kjønn, fødselsår, morsmål og fødselsmåned. Det første instrumentet i undersøkelsen hadde til hensikt å måle elevenes mestringsforventninger, og det er dette måleinstrumentet som er relevant for problemstillingen denne oppgaven baserer seg på. Del to av undersøkelsen besto av Usher og Pajares (2009) sitt validerte instrument for å måle kildene til mestringsforventninger. Del tre hadde til hensikt å måle mestring i matematikk, og besto av de samme oppgavene fra mestringsforventningsinstrumentet, men i ulik rekkefølge og skulle løses. Det vil være viktig å påpeke at spørreskjemaet og måleinstrumentet denne oppgaven baserer seg på ble utviklet av alle medlemmene i SESAM. Hver del har blitt nøye diskutert og gjennomgått, og derfor vil hele denne prosessen beskrives nedenfor, til tross for at ikke alle delene er relevant for min problemstilling. Det vil dog fokuseres i størst grad på instrumentet som måler mestringsforventninger.

Arbeidet med denne masteroppgaven ble innledet med en pilotstudie. Nedenfor vil jeg presentere utviklingen av spørreundersøkelsen og måleinstrumentene, med utgangspunkt i hvordan det ble brukt i pilot-studien. Det ble gjort få endringer fra denne til hovedundersøkelsen. Disse endringene beskrives til slutt i denne delen.

Med inspirasjon fra Chen (2003) valgte vi å ta i bruk items fra en allerede godkjent og validert matematikkprøve som utgangspunkt for både mestringsinstrumentet og mestringsforventningsinstrumentet. I spørreundersøkelser hvor det stilles spørsmål som krever en viss kunnskap fra respondentene, er det en forutsetning at respondentene faktisk har denne for å kunne besvare spørsmålene (Johannessen et al., 2012). På bakgrunn av dette valgte vi å ta i bruk items fra den internasjonale undersøkelsen TIMSS, som har til hensikt å måle elevers kompetanse i naturfag og matematikk (Utdanningsdirektoratet, 2016). TIMSS-studien gjennomføres i regi av International Association for the Evaluation of Education Achievement, og den første gjennomføringen av undersøkelsen fant sted i 1995. Siden da har den blitt gjennomført hvert fjerde år. Undersøkelsen er designet for å fange bredden til fagene slik de blir undervist i deltakerlandene, og baserer seg på informasjon om pensum og læreplaner, samt empirisk informasjon om konteksten for skolegangen i landene (IEA, u.d). Vi har tatt utgangspunkt i oppgavene fra TIMSS 2015. Hver gang TIMSS gjennomføres, blir

halvparten av oppgavene som brukes det året publisert. Det var disse oppgavene vi hadde til rådighet. Siden TIMSS-undersøkelsen gjennomføres på 9.trinn i Norge, valgte vi også å gjennomføre våre undersøkelser på dette trinnet. Spørreundersøkelsen ble gjennomført på papir, da dette ble ansett enklest mulig, i den forstand at ikke alle klasser har tilgang til datamaskin/læringsbrett i alle timer ved skolene vi skulle besøke.

3.4.1 Bakgrunnsvariabler

For å kunne besvare oppgavens problemstilling var det nødvendig å ha noe bakgrunnsinformasjon om respondentene. Derfor ble bakgrunnsvariablene kjønn, fødselsår, fødselsmåned og morsmål inkludert. Kjønn, fødselsår og morsmål blir i denne sammenheng dikotome variabler, som er variabler med kun to verdier. For variabelen kjønn var jente/gutt alternativene respondentene kunne velge mellom. For fødselsår og morsmål kunne de velge mellom 2006/ annet, og norsk/annet. Den siste bakgrunnsvariabelen er fødselsmåned og her blir elevene bedt om å krysse av for hvilken måned de er født i, og er dermed eksogene variabler på ordinalnivå.

3.4.2 Mestringsforventninger

Det er ifølge Pajares og Miller (1995) flere måter å måle mestringsforventninger. Den mest teoretiske passende og empiriske garanterte er om målingen av mestringsforventningene tar utgangspunkt i de samme eller lignende ferdigheter som kreves for oppgavene som skal løses. På bakgrunn av dette ble det valgt at delen som har til hensikt å måle elevenes mestringsforventning var de samme oppgavene som elevene ble bedt om å løse i undersøkelsens del om mestring. Oppgavene som utgjorde instrumentet for måling av mestring i matematikk, som presenteres nedenfor, var dermed også utgangspunktet for måling av mestringsforventninger. Bandura (1997) understreker at mestringsforventninger er kontekstspesifikt, og varierer avhengig av oppgavens vanskelighetsgrad og av ulike områder innenfor matematikken. For å kunne måle elevenes mestringsforventninger på best mulig måte, var det viktig at vår undersøkelse inneholdt oppgaver fra ulike emneområder og av ulik vanskelighetsgrad.

Oppgaver fra TIMSS 2015 ble valgt ut siden de var de nyeste oppgavene vi hadde til rådighet. TIMSS-undersøkelsen innenfor matematikk deles inn i fire emneområder; *tall, algebra, geometri og statistikk*. 24 oppgaver innenfor emneområdet tall, 26 oppgaver innenfor algebra,

19 oppgaver innenfor geometri og 16 oppgaver innenfor statistikk ble publisert fra TIMSS 2015. For å dekke matematikkfaget i størst mulig grad, ble det bestemt at instrumentet skulle inneholde like mange oppgaver fra hvert av emneområdene. Det ble også vektlagt at oppgavene innenfor hvert av emneområdene skulle ha ulik vanskelighetsgrad, da dette er en viktig del av måling av mestringsforventninger og for at alle elever skal ha mulighet til å mestre noen av oppgavene. Mange av oppgavene i TIMSS 2015 er flervalgsoppgaver, men det ble bestemt at vi ikke skulle inkludere disse oppgavene for at elevenes svar minst mulig skulle preges av gjetting. For å gjøre denne utvelgelsesprosessen enklest gjorde alle medlemmene i SESAM et utvalg, hver for seg, av oppgaver basert på emneområde og vanskelighetsgrad. Deretter la vi frem oppgavene vi hadde valgt ut, og ble enige om hvilke oppgaver som skulle utgjøre instrumentet som måler mestringsforventninger i matematikk hos elever på 9.trinn. Fire oppgaver fra hvert av emneområdene ble valgt ut, og disse oppgavene utgjorde delen av undersøkelsen som tok for seg mestringsforventninger og mestring i matematikk.

3.4.2 Kilder til mestringsforventninger

Denne delen av undersøkelsen hadde til hensikt å måle kildene til elevenes mestringsforventninger, og baserer seg på et instrument utviklet av Usher og Pajares (2009). Instrumentet er opprinnelig på engelsk, og vår hovedoppgave i denne delen var å oversette disse utsagnene slik at de beholdt sin opprinnelige mening. Alle i SESAM oversatte utsagnene hver for oss. Deretter satt vi sammen og diskuterte ulike formuleringer. I denne prosessen ble ordvalg spesielt vektlagt. Som nevnt over er det viktig at spørsmål i et spørreskjema er konkrete, entydig og lett forståelig for respondentene. Dermed var det viktig for oss å ta i bruk ord som elevene hadde et kjennskap til, og som brukes i dagligtalen. Vi valgte derfor å bruke ordet *matte*, istedenfor *matematikk*, da vi opplever at det er *matte* elevene bruker når de snakker om matematikkfaget i skolen. I tillegg valgte vi å oversette *problem* til *oppgave*, slik at det skulle være enklere for elevene å forstå spørsmålene. Utsagn fire «i got good grades on my last report card» ble situasjonstilpasset til det norske skolesystemet «jeg fikk god karakter i matte til sommeren i 8.klasse». Videre valgte vi å beholde ordet «pushe» også i de norske formuleringene. Vi fant ikke et tilsvarende norsk ord som omfavnet de samme assosiasjonene vi hadde med «pushe». I tillegg anså vi dette ordet som et ord som også brukes i dagligtalen i Norge. Utsagnet «adults in my family have told me what a good math student I am» ble omformulert til «voksne i familien min har fortalt meg at jeg er flink i matte», siden vi opplever at en i Norge sjelden snakker om at man er en «god elev». Fokuset for hele denne

prosessen var å gjøre utsagnene enkle for elevene å forstå, samtidig som innholdet eller det de formidler ble det samme som på engelsk. Dette er en prosess som er viktig i utformingen av et spørreskjema. Vi diskuterte også hvordan minoritetsspråklige elever ville forstå de ulike utsagnene. På bakgrunn av dette valgte vi også å ha med et spørsmål på første side om hvilket morsmål elevene hadde. Elevene fikk da valget mellom norsk eller annet. På denne måten kunne vi få innsyn i om en elev kunne ha forståelsesproblemer.

3.4.5 Mestring i matematikk

For å måle mestring, skulle elevene regne ut oppgave de hadde tatt stilling til i instrumentet som måler mestringsforventninger. Oppgavene kom i en annen rekkefølge enn i undersøkelsens første del, og elevene fikk i tillegg beskjed om at de ikke kunne gå tilbake å endre svarene fra undersøkelsens første del. Elevene fikk ikke tilgang på hjelpemidler, men alle fikk tilgang til kladdark.

3.5 Pilot-studie

Før instrumentene og undersøkelsen kunne ferdigstilles ble det gjennomført en pilot-studie. En pilotstudie gjennomføres for å bedre reliabiliteten, sikre at instrumentet måler det den skal og for å undersøke de praktiske sidene ved gjennomføringen (Johannessen et al., 2012). Et annet aspekt ved å gjennomføre en pilotundersøkelse er å teste hvordan måleinstrumentet fungerer i en Rasch-analyse, og dens krav om endimensjonalitet. En slik studie kan gjennomføres på forskere med erfaring med metoden, fagfolk som kjenner feltet eller personer som har like kjennetegn som de som skal undersøkes. SESAM valgte å involvere personer som hadde fellestrekk med de som skal delta i hovedstudien, og pilot-studien ble derfor gjennomført i klasser på 9.trinn. Alle elevene, utenom en elev, ble ferdig med undersøkelsen i løpet av 60 minutter. Vi fikk inn 45 besvarelser, men én respondent hadde enn annen alder enn resten. En respondent ble ikke ferdig med del tre, og to besvarelser besto kun av spørsmålsteget som svar på oppgavene i del tre. Derfor ble kun 41 besvarelser brukt i analysene. Dette er godt over anbefalt minstekrav på $N=30$ (Cohen et al., 2018).

Av de 41 besvarelsene vi fikk, hadde ni respondenter krysset av for *annet* som morsmål. Noen av disse besvarelsene hadde «ulogiske» svar på oppgavene. På oppgaven med «hvor mange elever likte historie best» var ett av svarene «sport» (se Vedlegg 1 for oppgavene). Det er vanskelig å vite om eleven skjønner teksten i matematikkoppgavene eller ikke. Om det er slik

at eleven får mange av oppgavene feil på grunn av språkferdigheter, måler man ikke mestring eller forventning om mestring i matematikk, men elevenes språkferdigheter. For å sikre at vi måler det vi ønsker å måle var det viktig å undersøke om elevene med annet morsmål påvirket validiteten og reliabiliteten i målingene. Om dette var tilfelle kunne det være hensiktsmessig å ekskludere elevene hvor språk kan spille en avgjørende faktor for om man greier å løse oppgaven eller ikke.

3.5.1 Analyse og resultat

Instrumentet som har til hensikt å måle mestringsforventninger er tuftet på instrumentet som måler mestring. Siden oppgavene fra TIMSS 2015 allerede er validerte matematikkoppgaver tilpasset elever på 9.trinn, valgte vi å la analysene for mestringsforventninger styre valget av involverte oppgaver i vår studie. På bakgrunn av at måling av mestringsforventninger er et instrument utviklet av SESAM ble det særlig viktig å sjekke at det testet det vi ønsket å teste, og derfor ble grundige analyser gjennomført på dette instrumentet.

41 besvarelser utgjorde pilot-studien, og Rasch-modellen ble brukt til å analysere dataene. Fit-analysene viste at det var gode fit-verdier (outfit og infit MNSQ 0,7-1,3) for alle items bortsett for item 3, 8, 11 og 15. Item 11 hadde høy MNSQ verdi for både infit (2,86) og outfit (2,00) og regnes da som underfit, og er en trussel mot endimensjonaliteten til måleinstrumentet. Også oppgave 8 hadde utilfredsstillende MNSQ-verdier <0,56. Oppgave 11 og 8 ble fjernet, og en ny analyse ble kjørt. Dette førte til at oppgave 15 ble mer problematisk med lave MNSQ-verdier. På bakgrunn av dette ble en ny analyse gjennomført, der item 8, 11 og 15 ble utelatt. Resultatene viste at item 3 fortsatt var problematisk, med noe høye MNSQ-verdier. På bakgrunn av at item 8, 11 og 15 er fra hvert sitt emneområde, henholdsvis algebra, geometri og statistikk, var det naturlig å også fjerne et item fra det siste emnet, tall. For å ha like mange oppgaver fra hvert emneområde ble det derfor besluttet å fjerne én oppgave fra emneområdet tall også. Siden item 3 ikke var fra dette emneområdet, ble det avgjort å fjerne en oppgave med tilnærmet samme vanskelighetsgrad. Konklusjonen ble å ta bort item 4, som bidrar på omtrent samme måte i testen. De samme analysene ble gjennomført på nytt, nå uten oppgave 4, 8, 11 og 15. Outfit MNSQ 0,79-1,31 og alle items er da innenfor grensen. Den standardiserte fit-statistikken (ZSTD) befinner seg også innenfor grensene -2 og 2. Analysene viser også Rasch reliabilitet estimat er 0,84 for personer og 0,95 for items. Disse verdiene indikerer reproducerbare tall, og at det er tilfredsstillende spredning av items og respondenter langs den ideelle linjen. Analysene viser også at det var samsvar,

med noen små forskyvninger, mellom forventet rangering av oppgavens vanskelighetsgrad og den rangeringen RSM gir oss. Det er også i tilfredsstillende grad korrespondanse mellom respondentenes mestringsforventninger og hvor vanskelig de ulike items oppfattes å være. ICC til noen tilfeldige item ble også sett på for å undersøke om det finnes en ekstra dimensjon i måleinstrumentet, og flesteparten holder seg innenfor 95% konfidensintervall mer eller mindre hele tiden.

De samme analysene ble også gjennomført, hvor elevene som hadde krysset av for «annet» morsmål, ble ekskludert. Dette utgjorde ingen stor forskjell på resultatene, og dermed ble det besluttet at alle elevene, uavhengig av morsmål, skulle inkluderes i hovedstudien. På bakgrunn av disse analysene ble fire oppgaver ekskludert, og de resterende 12 oppgavene utgjorde instrumentet som måler mestringsforventninger som vi brukte i hovedundersøkelsen (se vedlegg 1).

Gruppen gjennomførte også analyse på resultatene fra instrumentet som måler kilder til mestringsforventninger. Resultatet av denne analysen presenteres ikke i min masteroppgave, da jeg ikke benyttet meg av dette instrumentet i mitt arbeid.

3.6 Hovedundersøkelsen

Selve datainnsamlingen til hovedundersøkelsen ble gjennomført i løpet av høsten 2020. Medlemmene av SESAM tok kontakt med aktuelle skoler ved å sende forespørsel om deltakelse til rektor ved ulike skoler. Kontakten med skolene og gjennomføring av undersøkelsen ble gjort fortløpende, for å gjøre prosessen enklest mulig både for oss i SESAM og skolene. Vi var ansvarlig for selve gjennomføringen av undersøkelsen på alle skolene, men det var alltid en lærer med kjennskap til klassen til stede i klasserommet mens undersøkelsen foregikk. I begynnelsen av hver gjennomføring ble undersøkelsen beskrevet, hvor det var viktig å få frem at undersøkelsen var fullstendig anonym, og at det var frivillig å delta. Videre ble det gjennomgått hvordan elevene skulle fylle ut skjemaet ved å vise eksempler med prosjektor. I forkant av gjennomførelsen ble vi i SESAM enige om hvilken informasjon elevene skulle få, og at alle skulle forholde seg til denne informasjonen i gjennomførelsen (se Vedlegg 2). Dette ble gjort for å sikre at alle elevene hadde tilgang til lik informasjon, slik at dette ikke var en faktor som kunne ha innvirkning på resultatet. Ved at

slike detaljer foreligger kan datainnsamlingen i større grad standardiseres, og sannsynligheten for tilfeldige målefeil vil dermed reduseres (Kleven & Hjordemaal, 2018).

Elevene ble først bedt om å fylle ut informasjon om bakgrunnsvariablene, kjønn, alder, fødselsår, fødselsmåned og morsmål. I gjennomføring av undersøkelsen første del, måling av mestringsforventninger, rådet det enighet i SESAM om å følge Chen (2003) fremgangsmåte. Elevene fikk utdelt spørrearket på papir, men oppgavene de skulle ta stilling til ble vist med prosjektor. Elevene skulle vurdere hvor trygge de var på om de greide å løse ulike matematikkoppgaver, på en skala fra 1 til seks, hvor 1 var *veldig utrygg* og 6 var *veldig trygg*. Alle oppgave ble vist og lest høy i fellesskap i klasserommet, og elevene skrev svaret sitt på papir. Ved å gjøre det på denne måten fikk vi mulighet til å lese hver oppgave høyt, slik at alle elevene hadde størst mulig sjanse for å forstå hva oppgaven innebar. I tillegg kunne vi da kontrollere hvor lang tid elevene fikk på å avgjøre hvor trygg de var på at de greide å løse oppgaven eller ikke. Ved å gjøre det på denne måten fikk vi mulighet til kontrollere at elevene ikke får nok tid til å faktisk løse oppgaven. Mestringsforventninger er kun et mål på hva elevene tror de kommer til å få til, og ikke hva de faktisk får til. Hva de faktisk får til måles i undersøkelsens mestrings-del. Det ble forklart at vi skulle vise 12 matematikkoppgaver, og at elevene skulle vurdere hvor trygge de var på å løse oppgaven. Dette ble spesifisert ved et eksempel, hvor både oppgave og svartabell ble vist.

Det var kun i undersøkelsens første del at SESAM deltok aktivt. Resten av undersøkelsen løste elevene selv, i eget tempo, men innenfor én skoletime. Undersøkelsens andre del hadde til hensikt å måle kilder til mestringsforventninger, og besto av en oversatt utgave av Usher og Pajares (2009) sitt instrument. Her ble elevene presentert for 24 utsagn de skulle ta stilling til. Et eksempel på et slik utsagt er «mattelærerne mine har fortalt meg at jeg er flink i matte». Elevene ble bedt om å vurdere hvor godt disse påstandene stemmet for seg selv, på en skala fra én til seks, hvor én betyr *stemmer ikke i det hele tatt*, og seks betyr *stemmer veldig godt*. Når elevene var ferdig med denne delen gikk de rett videre til del tre, hvor de ble bedt om å løse matematikkoppgavene fra del én. Elevene hadde ingen tilgang til hjelpemidler, men fikk utdelt kladdark for utregninger.

3.6.1 Populasjon og utvalg

En undersøkelses målgruppe betegnes som studiens populasjon, og kan i mange tilfeller bestå av mange enheter (Johannessen et al., 2012). I studier der populasjonen er stor er det vanlig å

gjøre et utvalg som vil fungere som representanter for populasjonen. I denne studien er målgruppen elever på 9.trinn i grunnskolen i Norge. På grunn av tid og ressurser var det ikke mulig å gjennomføre undersøkelsen for hele målgruppen, og undersøkelsen ble derfor gjennomført på et utvalg av elever på 9.trinn. Fra denne store teoretiske populasjonen, ble det foretatt en utvelgelse basert på bekvemmelighet. Bekvemmelighetsutvalgt baserer seg på utvelgelse av de enheter som er lett tilgjengelig (Ringdal, 2013). Skolenes geografiske beliggenhet og om skolene hadde tid og mulighet var faktorer som påvirket utvelgelsesprosessen. I løpet av høsten 2020 ble det sendt ut mail til rektor eller fagansvarlig i matematikk ved 38 skoler med forespørsel om å kunne gjennomføre undersøkelsen på 9.trinn ved skolen. 13 skoler svarte ja til å delta i undersøkelsen. Det var endel skoler som svarte nei til denne gjennomførelsen, av ulike årsaker. Det vil i den sammenheng være viktig å nevne at denne undersøkelsen har blitt gjennomført i en tid med mange restriksjoner grunnet koronapandemien. Mange skoler, spesielt i Oslo, var underlagt strenge tiltak for å hindre smittespredning. Det var derfor ikke mulig å få gjennomført undersøkelsen ved disse skolene. Det vil i den forbindelse være viktig å påpeke at denne utvelgelsen ikke skyltes elevenes valg og motivasjon, heller ikke rektorers valg om å ikke delta i undersøkelsen. Kleven og Hjordemaal (2018) understreker at man må være forsiktig med å generalisere ut fra utvalg som ikke er basert på et sannsynlighetsutvalg. På bakgrunn av dette vil det ikke være mulig å kunne generalisere resultatene vi finner til å gjelde alle elever på 9.trinn i Norge.

Utvalget besto av 583 respondenter fra Møre og Romsdal, Viken, Agder og Oslo. Alle elever som var til stede, og fulgte læreplanen deltok i undersøkelsen. Et fåtall elever fulgte ikke i ordinær matematikkundervisning. Om disse elevene skulle delta eller ikke ble diskutert med faglærer, som mente at elevene det var snakk om ikke hadde forutsetning til å kunne utføre oppgavene i undersøkelsen siste del. Faglærer anbefalte at disse elevene fulgte eget opplegg, og deltok derfor ikke i denne undersøkelsen. For denne masteroppgaven er ikke alle enhetene i utvalget interessante for problemstillingen, og dermed er fokuset på to undergrupper fra utvalget, henholdsvis elever født tidlig og sent på året.

Det er stor variasjon i hvilken oppdeling studier baserer seg på når det kommer til inndeling av det av når elever er født. Frøseth et al. (2008) deler ikke inn elevene etter fødselsmåned for sammenligning, men fastslår en lineær alderseffekt innenfor hvert fødselsår. Gabrielsen og Lundetræ (2017) baserer seg på data fra PIRLS og tar for seg elever født i januar/februar og november/desember for å belyse relativ alderseffekt. Strøm (2004) tar i bruk data fra PISA-

undersøkelsen og sammenligner elever født i januar med elever født i desember. Dette er studier med tilgang til store datamateriale, og dermed har de mulighet til å undersøke ytterpunktene. Reed et al. (2017) som undersøker om RAE eksisterer i sport i skolen, deler året inn i fire deler, og sammenligner dermed elever født innenfor tre måneder i hver ende av det akademiske året. Lien et al. (2005) baserer seg på å dele året inn i tre deler, og sammenligner dermed elever født i januar, februar, mars og april mot september, oktober, november og desember. Johannessen et al. (2012) påpeker at viktige undergrupper bør representeres av minst 100 representanter. På bakgrunn av dette har jeg valgt, i likhet Lien et al. (2005), å dele året inn i tre, og dermed sammenligne elever født i årets første fire måneder med elever født de siste fire månedene. Dermed baserer mine analyser på svar fra 178 elever født i januar, februar, mars og april og 187 elever født i september, oktober, november og desember. Av undergruppen som var født i løpet av årets første fire måneder var 84 jenter og 94 gutter, og av elevene som var født i løpet av årets fire siste måneder var 91 jenter og 96 gutter.

3.6.2 Feilkilder

Det ble observert at noen elever gikk tilbake til del én og endret svaret sitt der. Det eleven(e) mest sannsynlig ikke tenkte over var at rekkefølgen var stokket om i de to ulike delene. Dette gjør at eleven har endret hvor trygg han var på å løse en oppgave på en oppgave som mest sannsynlig ikke stemte overens i de to delene. Dette kan være en feilkilde til datainnsamlingen. Det ble også reist spørsmål om hva som mentes med begrepet morsmål. Dette var et spørsmål vi hadde diskutert på forhånd, og ble besvart med «det språk du prater mest når du er hjemme». Også spørsmål om hva som menes med «gode karakterer» ble besvart med «det er din oppfatning som gjelder, det er hva du selv opplever som gode karakterer».

Det er ofte knyttet vanskelighetsgrad til å kartlegge personers holdninger og atferd. Observatøreffekten brukes ofte for å illustrere problemene knyttet til å få tak i hva som er typisk for en person (Kleven & Hjørdemaal, 2018). Denne effekten vil i stor eller liten grad ha en viss effekt på respondentene, og selv om den vil variere avhengig av situasjon må man alltid regne med at den er til stede. Et tilsvarende problem knyttet til målesituasjoner er når respondentene skal vurdere eller fortelle om seg selv. Slike situasjoner innebærer alltid en mulighet for å gi et mer positivt bilde av seg selv, og dermed fortelle like mye om hvordan

man ønsker å være som den man egentlig er (Kleven & Hjordemaal, 2018). I prinsippet må man alltid regne med at slike feilkilder oppstår i større eller mindre grad.

3.7 Statistiske analyser

Dataprogrammet WINDSTEPS 3.81.0 Software ble brukt til å validere instrumentet for mestringsforventninger og for å analysere dataene fra hovedstudien. Resultatet fra denne analysen vil videre bli brukt i en gjennomsnittsanalyse for å undersøke om det er sammenheng mellom RAE og mestringsforventninger i matematikk for elever på 9.trinn.

3.7.3 Gjennomsnittsanalyse

Gjennomsnittsanalyse er en analyse som gjennomføres for å undersøke forskjeller mellom grupper (Ringdal, 2013). En slik analyse kan svare på to spørsmål; *er det forskjell mellom gruppene? Gjelder disse forskjellene også i populasjonen?* Det siste spørsmålet forutsetter at utvalget er gjenstand for et tilfeldig utvalg. På bakgrunn av at populasjonen denne masteroppgaven baserer seg på ikke baserer seg på tilfeldig utvalg, vil det heller ikke være mulig å besvare dette spørsmålet. Jeg vil derfor kun fokusere på om det er forskjeller mellom gruppene i utvalget, og gjennomføre t-test for å undersøke dette. T-tester er basert på utregning av gjennomsnitt, og fungerer dermed bare for kontinuerlige variabler. Data fra Liker-skalaer er imidlertid på et ordinært målenivå, men på grunnlag av at vi har hentet gjennomsnittet for de to gruppene fra Rasch-modellen kan vi likevel bruke t-test. En forutsetning for å kunne ta i bruk t-test er at datamaterialet er normalfordelt.

For å besvare forskningsspørsmålene er det utviklet hypoteser, og resultatene fra t-testen vil bli brukt for å besvare disse. På bakgrunn av et mine hypoteser, som jeg har valgt å presentere i Resultat-kapittelet, består av både ensidige og tosidige hypoteser, vil det variere om det rapporteres enhalet-p eller tohalet-p. Dette spesifiserer i rapporteringen. For å undersøke styrken på sammenhengen mellom gruppene, rapporteres også Cohens d.

3.8 Etikk og personvern

All forskningsvirksomhet er underlagt juridiske retningslinjer og etiske prinsipper (Johannessen et al., 2012). Forholdet mellom mennesker, og hvordan vi direkte eller indirekte påvirker hverandre reiser mange etiske spørsmål. Alle forskere i Norge må følge de forskningsetiske retningslinjene vedtatt av den nasjonale forskningsetiske komité for

samfunnsvitenskap og humaniora. Disse retningslinjene går på beskyttelse av respondentene i et forskningsprosjekt (Ringdal, 2013). Retningslinjene stiller bestemte krav til forskningsprosessen for å sikre respondentene selvbestemmelse og autonomi, beskyttelse mot skade og respekt for respondentens privatliv. I følge Johannessen et al. (2012) utløses meldeplikt om forskningsprosjekter hvis prosjektet omfatter personopplysninger og disse opplysningene lagres elektronisk. En slik søknad sendes og behandles hos Norsk senter for Forskningsdata.

3.8.1 Søknad til Norsk senter for Forskningsdata (NSD)

Søknad til NSD ble sendt inn i slutten av september 2020. Det ble konkludert at prosjektet ikke behandler personopplysninger, og dermed ikke trengte en vurdering av NSD (se vedlegg 5)

3.8.2 Samtykke

I samtale med NSD ble det konkludert at undersøkelsen ikke behandler direkte eller indirekte opplysninger som kan virke identifiserende. Det ble også påpekt at om undersøkelsen er anonym, trenger man ikke å innhente samtykke fra andre enn respondenten som deltar i undersøkelsen (se vedlegg 5). På bakgrunn av dette ble det ikke sendt ut et samtykkeskjema til foresatte. Det ble likevel forfattet et informasjonsskriv, som ledelsen av skole kunne sende med elevene hjem (se vedlegg 3). Det ble understreket for elevene at det var frivillig å delta i undersøkelsen. Respondentene hadde full rett til å trekke seg helt frem til skjemaet ble levert. Det ble informert om at spørreundersøkelsen var anonym, og man ikke skulle skrive navn på noen av sidene. På grunn av denne anonymiseringen var det ikke mulig til å trekke seg i etterkant av selve datainnsamlingen, da det var umulig å spore ett svar tilbake til enkelteleven. Det ble også påpekt at resultatet kun skulle brukes i vårt forskningsprosjekt, og ikke ville ha noen innvirkning på karakterene deres i matematikk.

4. Resultater

For å kunne besvare forskningsspørsmålene vil dette resultatkapittelet være todelt. Første del av kapittelet har til hensikt å besvare det første forskningsspørsmålet om hvordan et instrument som har til hensikt å måle mestringsforventninger hos elever på 9.trinn vil kunne se ut. På bakgrunn av at vi ikke fant et instrument som måler elevens mestringsforventninger i matematikk i norsk kontekst, så vi oss nødt til å utvikle dette instrumentet selv. Nedenfor vil jeg derfor presentere valideringsprosessen til dette instrumentet basert på dataene samlet inn i hoveddata-innsamlingen (N=583). Disse analysene bygger på de analysene som ble foretatt på pilotdataene (N=41), og som ble presentert i metode-kapittelet. Av de tre måleinstrumentene som utgjorde spørreundersøkelsen vil jeg i denne delen kun ta for meg instrumentet som måler mestringsforventninger, i den forbindelse vil bruk av begrepet *instrument* kun henvise til dette. Denne første delen av resultatkapittelet legger grunnlaget for at de øvrige forskningsspørsmålene kan besvares i resultatkapittelets andre del.

4.1 Validering av måleinstrumentet

Den viktigste formen for validering av et måleinstrument er å undersøke om man måler det man ønsker å måle, nemlig konstruktvaliditet. Rasch-analyse kan gi oss mål på konstruktvaliditet, basert på den tekniske kvaliteten til de ulike items (Bjerke & Eriksen, 2016). For å validere instrumentene ble valideringsprosessen som brukes av Bjerke og Eriksen (2016) og Pampaka et al. (2012) som igjen følger retningslinjene fra Wolfe og Smith Jr (2007a) og Wolfe og Smith Jr (2007b) fulgt. Windstep 3.81.0 Software ble brukt å sjekke om det var samsvar mellom våre data og RSM, og nedenfor vil jeg presentere resultatene av de analysene som ble beskrevet i metodekapittelet.

Rådataene som er samlet inn gjennom de 12 polytome itemsene i instrumentet er ordinale, og avstanden mellom responskategorier er nødvendigvis ikke like. Dette er som nevnt over en av begrensningene ved bruk av Liker-skalaer i undersøkelser, og et skaleringsproblem som klassisk testteori ikke tar hensyn til. Rasch-analyse derimot tillater oss en transformasjon av de ordinale dataene til kontinuerlige data, og videre til en intervallskala (Logit-skalaen) som bevarer rekkefølgen på de ulike items, forutsatt at dataene tilfredsstillers modellens vilkår. RSM er velegnet for polytome Likert-skalaer, og er den modellen som tas i bruk her. For at et måleinstrument skal være valid og reliabelt stiller Rasch-modellen krav til at det skal være endimensjonalt og invariant. For å validere instrumentet er det viktig å avdekke eventuelle

avvik mellom forventet- og observert verdi for de ulike itemsene i instrumentet. Dette gjøres ved å sjekke endimensjonalitet ved hjelp av fit-statistikk.

4.1.1 Instrumentets reliabilitet

Rasch-modellens indekser har til hensikt å hjelpe forskeren å avgjøre om det er nok spredning av items langs den ideelle linjen, og nok spredning i respondentenes oppfattede evner til å kunne besvare disse. Denne statistikken blir rapportert ved hjelp av to målinger; personreliabilitet og itemreliabilitet. Analysene av dette instrumentet rapporterer om en personreliabilitet på 0,89 og itemreliabilitet på 0,99. Med andre ord kan det tolkes som et vi har utviklet en hendelseslinje med god spredning i respondenters opplevde evner og items oppfattede vanskelighetsgrad. Høye verdier for disse indikerer også reproduserbare tall. Bond og Fox (2015) påpeker likevel at høy oppnåelse av Rasch reliabilitets estimat ikke signaliserer enden på valideringen, men å ha tilfredsstilt disse kriteriene gir oss mulighet til å vurdere andre målinger for å kunne avgjøre om instrumentet kan regnes som valid.

4.1.2 Endimensjonalitet

At instrumentet er endimensjonalt innebærer at det kun måler ett teoretisk begrep, og ikke utilsiktet forveksler to eller flere variabler på samme tid. Gjennom fit-statistikken kan man avdekke avvik mellom observerte verdier (N=583) og verdiene som forventes av Rasch-modellen, og dermed undersøke hvor godt hver item passer inn i det underliggende konstruktet (mestringsforventninger). Rasch-modellens krav om endimensjonalitet holdes tilstrekkelig godt for dataene, da MNSQ fit-statistikken viser verdier innenfor akseptable grenser (0,6-1,4) for alle 12 items. Den standardiserte fit-statistikken (ZSTD) viser imidlertid uforutsigbarhet i dataene for item 1, 2, 9 og 12. Hvor for item 1, 9 og 12 rapporteres det om mindre variasjon enn RSM forventet, og for item 2 rapporteres det om mer variasjon. Linacre (u.d.) skriver derimot at så lenge MNSQ-verdiene er akseptable, kan man se bort fra små variasjoner i ZSTD.

Tabell 1 - Fit-statistikk for hele utvalget (N=583)

Item entry	Measure	Score	Standard Error	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PT Measure	
								CORR.	EXP
1	.76	1563	.04	.79	-4.0	.80	-3.5	.75	.71
2	.59	1663	.04	1.22	3.6	1.22	3.4	.66	.71
3	-.50	2226	.05	1.26	3.9	1.14	1.9	.69	.70
4	-1.07	2438	.05	1.05	.8	.89	-1.2	.70	.68
5	.16	1900	.04	1.11	1.9	1.08	1.3	.69	.71
6	-.16	2057	.04	1.02	.3	.99	-.1	.71	.71
7	-.48	2208	.05	.96	-.6	.89	-1.5	.73	.70
8	.46	1727	.04	1.02	.4	1.03	.5	.72	.71
9	.07	1945	.04	.84	-2.8	.86	-2.2	.73	.71
10	.56	1677	.04	.98	-.4	.98	-.3	.68	.71
11	-1.11	2449	.06	1.33	4.2	1.04	.5	.67	.68
12	.72	1574	.04	.79	-4.0	.80	-3.5	.74	.71

Items ble i utgangspunktet rangert av medlemmene i SESAM etter forventet og tilskattede vanskelighetsgrad for å kunne sammenligne rangeringen med de empiriske funnene. De ulike items ble kategorisert i tre grupper; lett, middels og vanskelig. Measure (se Tabell 1) viser hvert items oppfattede vanskelighetsgrad. Disse verdiene for populasjonen viser et hierarki av items som samsvarer omtrent med den rangeringen SESAM utviklet, sett bort fra item 1. Item 1 ble av oss kategorisert som en lett oppgave, men som av RSM estimeres til å være den oppgaven elevene har lavest forventning til å mestre, og som derfor oppfattes som den vanskeligste oppgaven. Det vil imidlertid være viktig å påpeke, som Bandura og Schunk (1981) også understreker, at i undersøkelser som har til hensikt å måle mestringsforventninger, er tiden elevene får til å vurdere vanskelighetsgrad svært kort. Ved første øyekast kan en oppgave fremstå som tilsynelatende vanskelig, men inneholder likevel ikke komplekse regneoperasjoner. Den delen av konstruksjonen mellom det itemet som elevene oppfattet som enklest og det itemet som ble oppfattet som det vanskeligste, er godt dekket, uten signifikant store avstander mellom hvert item (alle under 0.6 logits), noe som indikerer en jevn gradering av opplevd vanskelighetsgrad. Det er likevel verdt å bemerke at elevene som rapporterer om høyest mestringsforventninger er lokalisert over de items som oppfattes som vanskeligst. Nye items er nødvendig for å kunne måle denne elevgruppen mer nøyaktig.

Neste steg i valideringsprosessen er å sjekke «Item polarity» som også er et mål for å vurdere instrumentets endimensjonalitet. Denne statistikken forteller om responsen på et gitt item stemmer overens med respondentenes opplevelse av egne evner og itemets opplevde

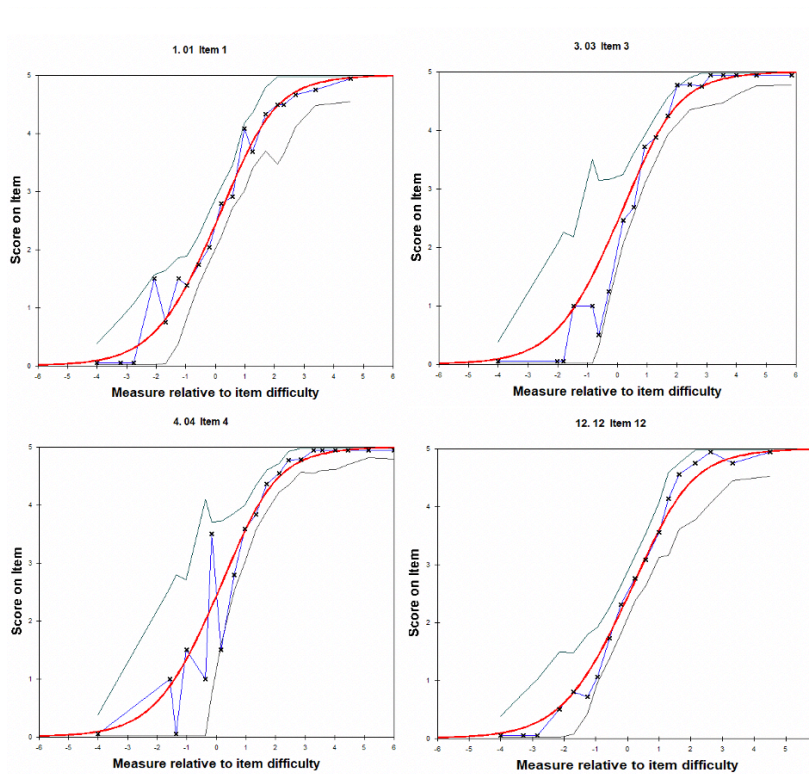
vanskelighetsgrad. Korrelasjonsverdier som er negative eller nær null signaliserer problematiske items som ikke er konsistent med det konstruert som måles, her mestringsforventninger (Bond & Fox, 2015). «Item polarity» rapporteres i PT-MEASURE CORR kolonnen i fit-statistikken (Tabell 1), og forventes å være positiv. Alle verdier ligger mellom 0,66-0,75, noe som tilsier at alle items følger retningen til den latente variabelen. RSM forventede verdi 0,68-0,71(PT-MEASURE EXP), og dermed er det tilfredsstillende korrespondanse mellom målt verdi og Rasch-modellens forventede verdi.

Det var heller ingen forstyrrelser av respondentenes gjennomsnittlige evner etter hver svarkategori for hvert enkel item (se Tabell 2). Dette betyr at svarkategoriene kom i stigende rekkefølge, fra veldig utrygg til veldig trygg. Ved bruk av Likert-skala i undersøkelser er det viktig å nettopp holde øye med rangeringen av svarkategoriene, slik at de opptrer slik de er tiltenkt. Hvis denne rekkefølgen ikke hadde vært stigende, indikerer det at noe er galt i testen, og videre analyser måtte ha blitt gjennomført.

Tabell 2 - Kategoristruktur basert på hele utvalget (N=583). Her vist for item 2

Category (item 2)	Observed		Ability Mean	Infit MNSQ	Outfit MNSQ
	Count	%			
0 Veldig utrygg	48	8	-1.34	1.1	1.1
1	65	11	.35	1.3	1.3
2	104	18	.58	1.2	1.3
3	147	25	.94	1.2	1.1
4	136	23	1.51	1.5	1.5
5 Veldig trygg	81	14	2.30	1.1	1.1

Videre sjekkes ICC for hvert enkelt item, som viser Rasch-modellens forutsigelse for mål relatert til opplevd vanskelighetsgrad for alle items. ICC for alle items ser omtrent lik ut (se Figur 2), og holder seg innenfor 95% konfidensintervall mer eller mindre hele tiden. Dermed viser analysene så langt ingen tegn til at instrumentet inneholder flere dimensjoner.



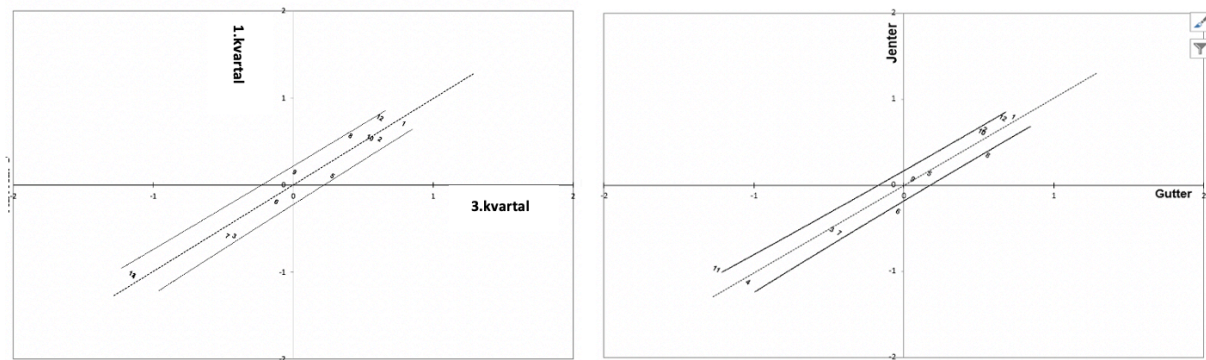
Figur 2 - Item Characteristics Curves for item 1, 3, 4 og 12

4.1.3 Invarians

Rasch-modellen legemliggjør et ideal som alltid «krenkes» av virkelig data (Linacre, 2012). Mange av disse bruddene er uvesentlig, uavhengig hvor statistisk signifikant de er. Men noen av bruddene har alvorlige konsekvenser, og må derfor tas på alvor. Hvis et instrument konsekvent fungerer eller oppfattes annerledes for en gruppe av respondentene enn for en annen, kan ikke instrumentet regnes som invariant. Det er to ulike måter å teste dette på, enten ved DIF (Differential Item Functioning) og DTF (Differential Test Functioning). I motsetning til DIF, som sjekker alle items hver for seg, undersøker DTF om hele instrumentet, bestående av alle tolv items, fungerer på samme måte for alle undergrupper i utvalget. Dette er en viktig del av valideringsprosessen til måleinstrumentet, for å sikre at studien ikke er preget av systematiske skjevheter (se kapittel 3.2.2 i metodekapittelet for mer detaljert beskrivelse).

To DTF-analyser ble utført for å bekrefte invarians av instrumentet, og kjørt på to undergrupper i utvalget. Hver DTF i Figur 3 viser de 12 itemsene i instrumentet. Ideelt sett bør alle punktene falle på den diagonale linjen, men i praksis blir instrumentet ansett som tilstrekkelig invariant hvis punktene er innenfor 95% konfidensintervall. DTF-analysen indikerer at alle 12 items ble oppfattet tilfredsstillende likt for elever født i 1.kvartal og

3.kvartal, og mellom jenter og gutter. Med andre ord utvises det ikke DTF i vårt datamateriale. Likevel ser vi at item 11 og item 8 ligger akkurat utenfor konfidensintervallet i analysen kjørt mellom jenter og gutter. Item 8 ligger rett utenfor til høyre, og det betyr at dette itemet oppleves vanskeligere for guttene enn RSM forventet, sett i forhold til jentene. Tilsvarende for item 11, som viser seg å oppleves vanskeligere for jentene enn RSM forventet. De ulike items ligger imidlertid like utenfor, og det gjøres derfor ikke noe med det. Det er likevel viktig å notere seg disse med tanke på konklusjoner som kan bli dratt på bakgrunn av disse items senere.



Figur 3 - Differential Test Functioning for elever født i 1. og 3. kvartal, og for jenter og gutter.

4.1.4 Oppsummering av valideringsprosessen

Basert på de ulike Rasch-analysene presentert over er det grunnlag for å konkludere med at instrumentet som måler mestringsforventninger i matematikk for elever på 9. trinn ikke inneholder en ekstra dimensjon, og at det ikke oppleves annerledes for de ulike undergruppene i utvalget. Dermed er instrumentet validert, og kan brukes for å måle norske elevers forventning om mestring i matematikk på 9. trinn.

4.2 Øvrige forskningsspørsmål

I denne delen av resultatkapittelet vil jeg presentere resultatene fra de statistiske analysene som er gjennomført på datamaterialet. Forskningsspørsmålene besvares ved bruk av deskriptiv statistikk og uavhengige t-tester. Deskriptiv statistikk omfatter de prinsipper som gir systematiske beskrivelser av datamaterialet som sentraltendens og spredning for variablene (Foldnes, Grønneberg & Hermansen, 2018). Det typiske målet for sentraltendens er gjennomsnittet av verdiene, som forteller oss om tyngdepunktet av verdiene i

datamaterialet. For å undersøke spredning, oppgis ofte varians (s^2) og standardavvik (SD). Gjennomsnittsskåren for elevenes mestringsforventninger er i denne masteroppgaven basert på Rasch-analyser.

Forutsetningen for å kunne si at eldre elever har høyere mestringsforventninger enn yngre elever, er å undersøke om det er forskjell mellom gruppene og om denne forskjellen er statistisk signifikant. For å vurdere dette benyttes en uavhengig t-test. En t-test forutsetter at variablene er kontinuerlige, og at datamaterialet er tilnærmet normalfordelt (Cohen et al., 2018). Statistisk signifikans tilsier bare å fortelle forskeren om et bestemt resultat har eller ikke har skjedd ved en tilfeldighet. I tillegg til statistisk signifikans er det også relevant å se styrken på denne sammenhengen, altså effektstørrelse. For å vurdere denne styrken vil jeg benytte meg av Cohens d , som ofte brukes i forskning hvor man sammenligner grupper med ulikt antall. Verdier for bruk av Cohens d presenteres nedenfor i Tabell 3. Estimering av statistisk styrke er viktig for at å betrakte sannsynligheten for å begå en Type I-feil og Type II-feil. En type I-feil er en falsk positiv, altså å avvise nullhypotesen når den i virkeligheten er sann. En type II-feil er en falsk negativ, altså å akseptere nullhypotesen når den i virkeligheten er falsk. Dataprogrammet SPSS brukes til å analysere datamateriale. Alle analysene blir presentert i tabeller nedenfor.

Tabell 3 - Verdier for Cohens d . Kilde: Cohen et al. (2018, s. 746)

Choens d	Effekt
0-0.20	Svak effekt
0.21-0.50	Liten effekt
0.51-1.00	Moderat effekt
>1.00	Sterk effekt

4.2.1 Relativ alder og mestringsforventninger

I denne delen presenteres resultatene fra analysene som er gjennomført for å undersøke om relativt eldre elever har høyere mestringsforventninger i matematikk enn relativt yngre elever, født innenfor samme skoleår. Denne hypotesen baserer seg på tidligere observerte korrelasjoner mellom RAE og prestasjon i matematikk (Aune et al., 2018; Bjerke et al., til fagfelle vurdering; Olsen & Bjørnsson, 2018), og mellom mestringsforventninger og

prestasjon i matematikk (Pajares & Kranzler, 1995; Pajares & Miller, 1995; Street et al., 2017). Følgende hypotese ble formulert:

De eldste elevene har høyere mestringsforventninger i matematikk enn de yngste elever

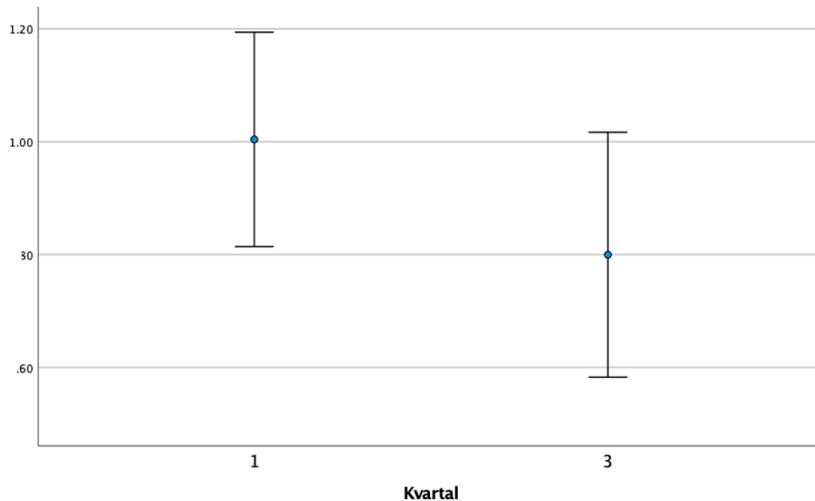
På grunn av ordlyden i hypotesen ser vi at hypotesen er ensidig, og derfor rapporteres det kun enhalet p-verdi.

Tabell 4 - Mestringsforventninger for elever født i 1. og 3.kvartal

	N	M	SD	MD	p	Choens d
1.kvartal	178	1.004	1.280			
				.205	.083*	0.146
3.kvartal	187	.799	1.504			

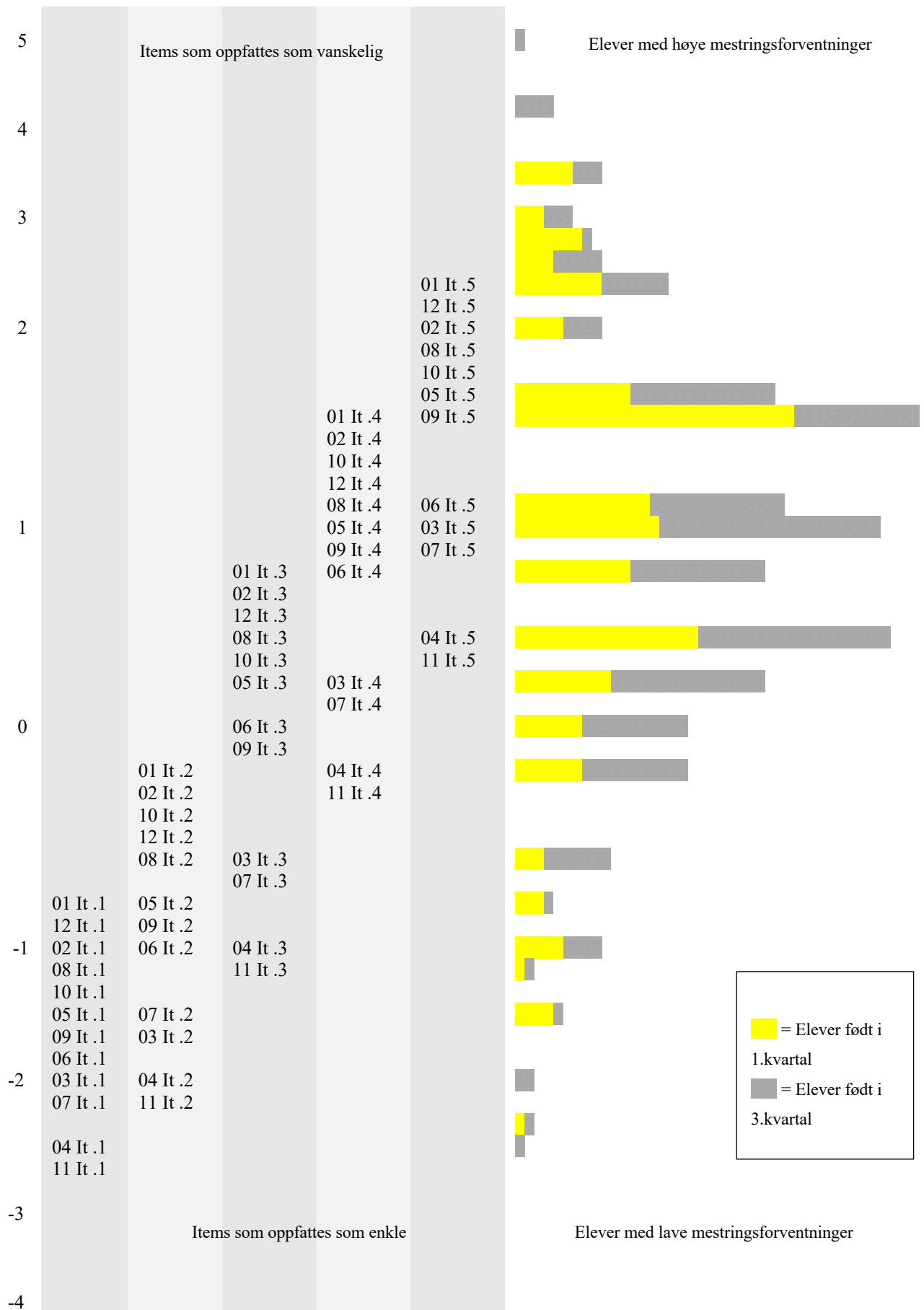
(*: signifikant på 0.1 nivå, **: signifikant på 0.05 nivå)

Tabellen ovenfor indikerer at elevene som er født i 1.kvartal (M=1.004, SD=1.280) hadde høyere mestringsforventninger enn elevene som er født i 3.kvartal (M=.799, SD=1.504), med en gjennomsnittlig forskjell på .205, og at denne forskjellen var signifikant ($t=1.392$, $df=362$, enhalet $p=0.0825$) på 0.1 nivå (men ikke på 0.5 nivå). På bakgrunn av at $p<0.1$ kan nullhypotesen, om det det ikke er forskjell mellom de eldste og yngste elevene, forkastes. Med andre ord viser resultatene fra analysene støtte for min hypotese om at de eldste elevene har høyere mestringsforventninger i matematikk enn de yngste elevene. Cohens d utviser imidlertid lav effektstørrelse ($d=.146$), noe som tyder på at forskjellen mellom gruppene er reell, men liten. Hypotesetesten bekreftet min hypotese, men jeg vil likevel grave dypere og finne ut mer om sammenhengen mellom i mestringsforventning i matematikk knyttet til de eldste og yngste elevene. Dette gjøres ved å undersøke gruppenes spredning, samt å fremstille elevenes skår ved hjelp av variabelkart. Dette illustreres nedenfor i Figur 5 og Figur 6.



Figur 4 - Gjennomsnitt og spredning i mestringsforventninger for elever født i 1. og 3. kvartal

Forskjellen mellom gruppene blir mer tydelig når man ser på spredningen i resultatet, hvor elever som er født i 3.kvartal har vesentlig høyere spredning, sammenlignet med elevene født i 1.kvartal. Figur 5 viser at resultatene til elevene født i 1.kvartal ligger mer samlet rundt gruppegjennomsnittet ($M=1.004$), sammenlignet med resultatene til elevene født i 3.kvartal ($M=0.799$). De yngre elevene ($SD=1.504$) utviser dermed større variasjon i vurderingen av egen forventning om mestring i matematikk, sammenlignet med de eldre elevene ($SD=1.280$). Videre undersøkes variabelkartet som viser rangeringen til de ulike items og elever. Figur 6 viser fordelingen av item og elever i variabelkartet. Det er imidlertid en skjevstilling i materialet, hvor de elevene med høyest rapportert forventning om mestring er lokalisert over de items som var oppfattet som vanskeligst. Nye items som kan gjenspeile disse elevenes oppfattede kompetanse er nødvendig for en mer nøyaktig estimering av elevene med høy mestringsforventning.



Figur 5 - Variabelkart som viser fordelingen av items og elever født i 1. og 3.kvartal

Forskning gjort av Bergem et al. (2016) viser at det er stor variasjon i hvor godt norske elever mestrer oppgavene innenfor hvert av emneområdene i TIMSS. For å undersøke om dette også er tilfelle med mestringsforventninger og RAE ble det undersøkt om elevene født i 1.kvartal (N_1) rapporterte om høyere mestringsforventninger sammenlignet med elevene født i 3.kvartal (N_3), knyttet til de tolv items, sett i lys av de fire emneområdene TIMSS baserer seg på. Dette belyses ved å legge frem deskriptiv statistikk knyttet til elevenes gjennomsnittlige skår item for item knyttet til emneområdene; *tall*, *geometri*, *algebra* og *statistikk*.

Tabell 5 viser gjennomsnittlig skår på hvert item på instrumentet som måler mestringsforventninger i matematikk til alle elevene som deltok i undersøkelsen ($N=583$), elevene født i 1.kvartal ($N_1=178$) og elevene født i 3.kvartal ($N_3=187$) knyttet til emneområdet tall, algebra, geometri og statistikk. De ulike items er rangert etter opplevd vanskelighetsgrad, hvor item 1 ble av elevene i hovedutvalget oppfattet som det vanskeligste, mens oppgave 11 ble oppfattet som det enkleste. Items som er farget blå tilhører emneområde *tall*, gule items hører til *algebra*, røde items tilhører *geometri* og de items som er farget grønn er innenfor emnet *statistikk*. Både item 1 og item 11, som ble oppfattet som det vanskeligste og det enkleste å løse, er innenfor emneområdet tall. Tre av de seks oppfattet vanskeligste items tilhører emneområdet algebra. Videre er det god spredning i andre items knyttet til emneområdet og opplevd vanskelighetsgrad. Tabellen viser også differansen til rapportert mestringsforventninger mellom elever født tidlig og elever født sent på året. Skåren knyttet til mestringsforventninger for hvert kvartal varierer, hvor elevene født i 1.kvartal hadde høyere forventning om mestring for item 1, 2, 5, 6, 7 og 3 enn elevene født i 3.kvartal. Elevene hadde lik forventning om mestring for item 10, og elevene født i 3.kvartal hadde høyere mestringsforventninger for de resterende fem items. Med andre ord rapporterer elever født i 1.kvartal høyere forventning om mestring for alle items innenfor emneområdet geometri. Elever født i 3.kvartal har høyere eller like høye mestringsforventninger for alle items innenfor statistikk. Forskjellen mellom elevene er imidlertid små og ikke signifikante.

Tabell 5 - Rangering av oppfattet vanskelighetsgrad knyttet til de ulike items. Item merket blå er fra emneområdet tall, gul=algebra, rød=geometri og grønn=statistikk.

Item	Measure N	Measure N ₁	Measure N ₃	Differanse N ₁ -N ₃
1	.76	.79	.71	.08
12	.72	.62	.78	-.16
2	.59	.62	.53	.09
10	.56	.55	.55	0
8	.46	.41	.57	-.16
5	.16	.28	.11	.17
9	.07	.01	.15	-.14
6	-.16	-.12	-.19	.07
7	-.48	-.47	-.58	.11
3	-.50	-.42	-.58	.16
4	-1.07	-1.14	-1.04	-.1
11	-1.11	-1.15	-.02	-.13

4.2.2 Mestringsforventninger, RAE og kjønn

For å undersøke om det er kjønnsforskjeller når det kommer til mestringsforventninger og RAE vil uavhengig t-test bli brukt for å sammenligne gruppens gjennomsnittlige resultat. Hypotesene som er formulert for å undersøke sammenhengen mellom relativ alder, mestringsforventninger og kjønn baserer seg på forskning gjort av Pajares (2005) og Jensen og Nordtvedt (2013) som viser at gutter uttrykker sterkere mestringsforventninger i matematikk enn jenter. På en annen side fant hverken Chen (2003) eller Pajares og Kranzler (1995) kjønnsforskjeller i mestringsforventninger i matematikk. Videre viser forskning at RAE gjennomgående er større for gutter enn for jenter (Solli, 2017). På bakgrunn av dette er det vanskelig å avgjøre i hvilken retning de eventuelle forskjellene vil vise seg, og det er dermed formulert tosidige hypoteser. For å undersøke om det er sammenheng mellom RAE, mestringsforventninger og kjønn er det gjennomført flere t-tester, med ulike kombinasjoner. Hypotesene knyttet til hver t-test presenteres nedenfor. På bakgrunn av at hypotesene er tosidige, rapporteres det tohalet p-verdi. Resultatene presenteres i tabeller nedenfor.

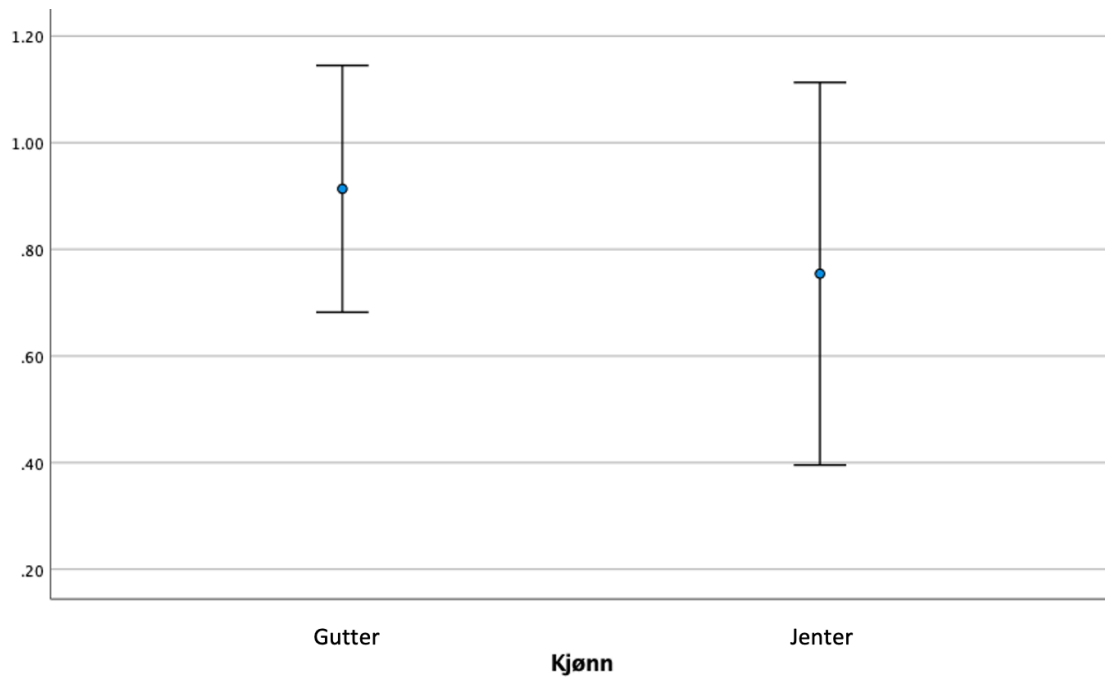
Som nevnt over er forskningen sprikende med tanke på om det er kjønnsforskjeller i mestringsforventninger i matematikk. På bakgrunn av dette er den første hypotesen formulert for å undersøke sammenhengen mellom gutter og jenter født i samme kvartal, og lyder slik:

Det er kjønnsforskjeller i mestringsforventninger i matematikk innad i hvert kvartal.

Tabell 6 indikerer at gutter og jenter født i 1.kvartal skårer tilnærmet likt (-.009) når mestringsforventninger i matematikk blir målt, men at gjennomsnittlig skår for guttene (M=1.014, SD=1.281) var noe lavere enn for jentene født i samme kvartal (M=1.023, SD=1.330). Forskjellen var derimot ikke signifikant ($t=-.046$, $df=176$, $p=.9642$), og nullhypotesen må derfor opprettholdes. Dette understrekes også av den lave effektstørrelsen ($d=-.007$). De relativt eldre jentene hadde imidlertid noe større spredning i sine resultater (SD=1.330), sammenlignet med de relativt eldre guttene (SD=1.281). I undersøkelsen av gutter og jenter født i 3.kvartal viser tabellen at relativt yngre gutter (M=.914, SD=1.142) rapporterer noe høyere mestringsforventninger sammenlignet med jenter født i samme kvartal (M=.754, SD=1.754), med en forskjell i gjennomsnitt på .159, altså motsatt av tendensen vi ser mellom kjønnene i 1.kvartal. Forskjellen mellom disse gruppene er heller ikke signifikante ($t=-.741$, $df=155$, $p=.460$, $d=-.11$). I likhet med de relativt eldre jentene hadde også de relativt yngre jentene større spredning i sine resultater, sammenlignet med guttene. Dette illustreres nedenfor ved Figur 7.

Tabell 6 - Mestringsforventninger knyttet til kjønn født i samme kvartal

	N	M	SD	MD	p	Choens d
1.kvartal G	94	1.014	1.281			
				-.009	.964	-.007
1.kvartal J	84	1.023	1.330			
3.kvartal G	96	.914	1.142			
				.159	.920	-.110
3.kvartal J	91	.754	1.754			



Figur 6 - Gjennomsnitt og spredning i mestringsforventninger mellom gutter og jenter født i 3.kvartal

Figur 7 viser forskjellen mellom rapportert mestringsforventning i matematikk mellom jenter og gutter født i 3.kvartal. Spredningen i jentenes resultat ($SD=1.754$) er vesentlig større sammenlignet med guttenes resultat ($SD=1.142$). Til nå har kjønnsforskjeller innad i hvert kvartal blitt undersøkt. Nedenfor vil jeg undersøke om det er kjønnsforskjeller mellom elevene født i ulike kvartal.

Videre i denne delen er det gjennomført fire ulike hypotesetester for å undersøke kjønnsforskjeller og RAE for mestringsforventninger i matematikk. Gallagher og Kaufman (2005) påpeker at forskning innen matematikk og kjønn i stor grad fokuserer på forskjeller mellom kjønnene, til tross for forskjellene innad i kjønnene sannsynligvis er større. På bakgrunn av dette undersøkes det dermed både forskjeller mellom kjønnene, men også forskjeller innad i kjønnene. Hypotesen er basert og formulert ut fra teori nevnt ovenfor. Nedenfor presenteres hypotesen som er utgangspunktet for testene. For å kunne undersøke hypotesen var det nødvendig å gjennomføre fire ulike hypotesetester, med ulike kombinasjoner med kjønn og kvartal. Hypotesen er tosidig, og det rapporteres derfor tohalet-p.

Det er forskjeller i mestringsforventninger i matematikk innad og mellom kjønnene for elever født i ulike kvartal.

Tabell 7 viser at mestringsforventningsskåren var høyere for gutter født i 1.kvartal ($M=1.014$, $SD=1.281$) enn for gutter født i 3.kvartal ($M=.914$, $SD=1.142$), men forskjellene er ikke signifikante ($t=.570$, $df=188$, $p=.2845$). Dette understreker også Cohens d , som viser en effektstørrelse på $.083$. Tabellen viser også at gutter født i 1.kvartal ($M=1.0139$, $SD=1.2813$) har høyere forventning om mestring sammenlignet med jenter født i 3.kvartal ($M=.7544$, $SD=1.7227$), men heller ikke her er forskjellene mellom gruppene signifikante ($t=1.165$, $df=183$, $p=.1225$). Cohens d viser derimot en svak effektstørrelse ($d=.171$), som kan indikere at det kan være en svak sammenheng mellom gruppene. Signifikansnivået ligger derimot godt over det som aksepteres innenfor samfunnsvitenskapelig forskning, og nullhypotesen må derfor opprettholdes. Tilsvarende ser vi at jenter født i 1.kvartal ($M=1.0229$, $SD=1.3302$) har høyere mestringsforventninger enn gutter født i 3.kvartal ($M=.912$, $SD=1.142$). Forskjellen mellom disse gruppene er heller ikke statistisk signifikant, og nullhypotesen opprettholdes også for denne kombinasjonen. Videre kan det se ut som at jenter født i 1.kvartal ($M=1.0229$, $SD=1.3302$) har høyere mestringsforventninger sammenlignet med jenter født i 3.kvartal ($M=.7544$, $SD=1.7226$). Selv om forskjellen i gjennomsnitt er $.2684$ er ikke denne signifikant ($t=1.147$, $df=173$, $p=.1265$). Cohens d utviser imidlertid også her en svak effektstørrelse ($d=.174$) noe som kan tyde på at det er en liten reell forskjell mellom gruppene. Siden $p>0.05$ beholdes nullhypotesen.

Tabell 7 - Mestringsforventninger knyttet til kjønn født i ulike kvartal

	N	M	SD	MD	p	Cohens d
1.kvartal G	94	1.014	1.281			
				.1003	.570	.083
3.kvartal G	96	.914	1.142			
1.kvartal G	94	1.0139	1.2813			
				.2595	.245	.171
3.kvartal J	91	.7544	1.7227			
1.kvartal J	84	1.0229				
			1.5208	.1089	.554	.089
3.kvartal G	96	.914				
1.kvartal J	84	1.0229	1.3302			
				.2684	.245	.174
3.kvartal J	91	.7544	1.7226			

4.2.3 Høyt-presterende- og lavt-presterende elever, mestringsforventninger og relativ alder

I denne delen av oppgaven presteres analysene, og resultatet av disse, for å undersøke om sammenhengen mellom fødselsmåned og mestringsforventninger opptrer forskjellig hos høyt-presterende elever enn hos lavt-presterende elever. For å undersøke dette er det gjennomført flere uavhengige t-tester. Hypotesene knyttet til hver test presenteres nedenfor. Forskning gjort av Aune et al. (2018) viste at jenter født sent på året var overrepresentert blant de lavt-presterende elevene, og at gutter født tidlig var overrepresentert blant høyt-presterende elevene. Dette undersøkes også i denne oppgave. Kategoriseringen av høyt-presterende og lavt-presterende elever er basert på undersøkelsens mestringsdel. Elevene som er kategorisert som høyt-presterende elever var de som fikk 7-12 poeng, og elevene som er kategorisert som lavt-presterende oppnådde 1-2 poeng. Nedenfor, i Tabell 8, presenteres fordelingen av jenter og gutter innenfor hver kategori. Denne viser at det er tilnærmet like mange jenter og gutter som kategoriseres som høyt-presterende og lavt-presterende i hvert kvartal.

Tabell 8 - Fordeling av høyt-presterende og lavt-presterende jenter og gutter født i 1. og 3.kvartal

	Høyt-presterende jenter	Høyt-presterende gutter	Lavt-presterende jenter	Lavt-presterende gutter
1.kvartal	20	19	17	14
3.kvartal	18	20	17	17

Collins (1982) fant at elever med høye mestringsforventninger var raskere til å avvise feil løsningsstrategi og løste flere oppgaver enn sine medelever som hadde lave forventninger om mestring, uavhengig om elevene var regnet som høyt-presterende eller lavt-presterende. Videre skriver Bandura (1977) at mestringserfaringer er den viktigste kilden til mestringsforventninger. Undersøkelser gjort av Jussim et al. (2009) viser også at lærere er mer oppmuntrende, viser mer støtte og gir mer oppmerksomhet til elever som forventes å prestere godt. Dette legger grunnlaget for følgende hypotese:

Høyt-presterende elever har høyere mestringsforventninger i matematikk enn lavt-presterende elever

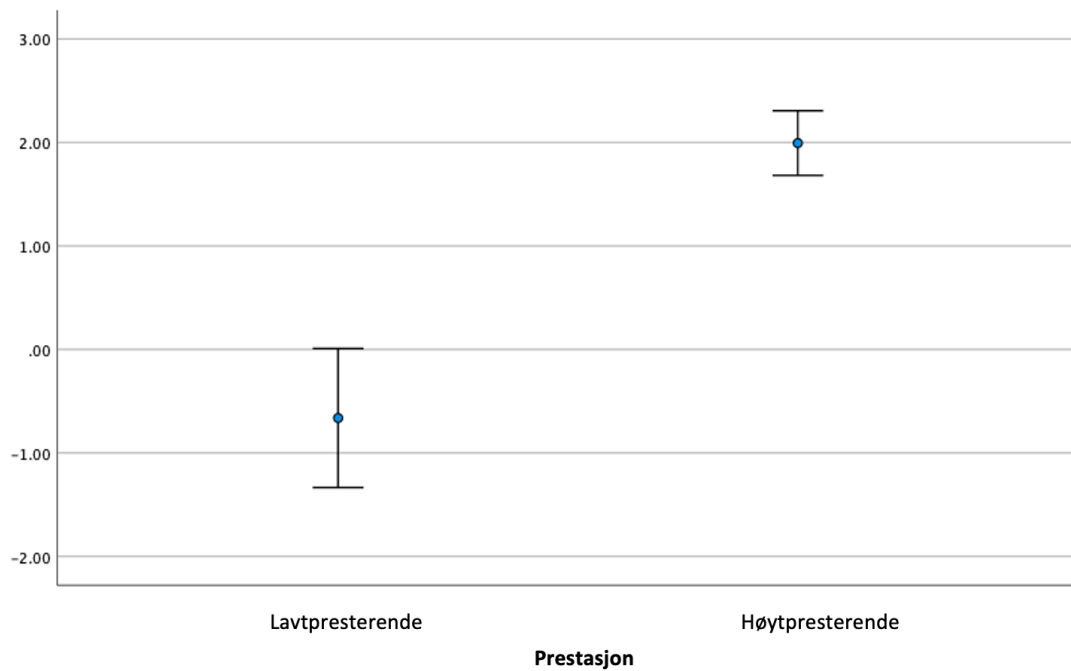
Basert på hypotesens ordlyd er hypotesen ensidig, og det rapporteres derfor enhalet p-verdi.

Tabell 9 viser at det var signifikant forskjell ($t=8.411$, $df=68$, $p<0.001$) i mestringsforventninger mellom høyt-presterende ($M=1.973$, $SD=.789$) og lavt-presterende ($M=.0268$, $SD=1.144$) elever født i 1.kvartal. Forskjellen i mestringsforventninger mellom høyt-presterende ($M=1.994$, $SD=0.959$) og lavt-presterende ($M=-.662$, $SD=1.924$) elever født i 3.kvartal var også signifikante ($t=7.293$, $df=46.986$, $p<0.001$). Videre ble også forskjellen mellom høyt-presterende elever født i 1.kvartal ($M=1.973$, $SD=.789$) og lavt-presterende elever født i 3.kvartal undersøkt ($M=-.662$, $SD=1.924$). Også denne forskjellen var statistisk signifikant. Cohens d viser også sterk effekt mellom alle gruppene. På bakgrunn av $p<.001$ kan nullhypotesen, om at det ikke er forskjell i gruppenes mestringsforventninger, forkastes. I videre undersøkelser ser vi også at lavt-presterende elever har vesentlig større spredning i sine resultater enn de høyt-presterende elevene. Denne forskjellen vises nedenfor, i Figur 8.

Tabell 9 – Mestringsforventninger til høyt-presterende og lavt-presterende elever

	N	M	SD	MD	p	Cohens d
1.kvartal H	39	1.973	.789			
				1.946	.000**	2.024
1.kvartal L	31	.027	1.144			
3.kvartal H	38	1.994	0.959			
				2.656	.000**	1.782
3.kvartal L	34	-.662	1.924			
1.kvartal H	39	1.973	.789			
				2.635	.000**	1.679
3.kvartal L	34	-.662	1.144			

(*:signifikant på 0.1 nivå, **:signifikant på 0.05 nivå)



Figur 7 - Gjennomsnitt og spredning i mestringsforventninger mellom høyt-presterende og lavt-presterende elever

Videre viser forskning innenfor RAE at elever født tidlig på året presterer bedre innenfor matematikk enn elever født sent på året (Bjerke et al., til fagfelleevaluering). Som nevnt over viser også forskning at det er sterk sammenheng mellom mestringsforventninger og prestasjon i matematikk (Street et al., 2017). På bakgrunn av dette er det interessant å undersøke om det er forskjell i mestringsforventninger mellom høyt-presterende elever født i ulike kvartal, og mellom lavt-presterende elever født i ulike kvartal. Med teorien nevnt over er det formulert en ensidig hypotese.

Eldre høyt-presterende elever har høyere mestringsforventninger i matematikk enn yngre høyt-presterende elever

Tabell 10 viser at gjennomsnittlig mestringsforventningsskår er tilnærmet likt for høyt-presterende elever født i 1.kvartal ($M=1.973$, $SD=.789$) og høyt-presterende elever født i 3.kvartal ($M=1.994$, $SD=.950$). Det er derimot en liten forskjell, hvor elevene som var født i 3.kvartal hadde noe høyere mestringsforventning enn de relativt eldre elevene ($MD=-.021$). Forskjellene mellom de to gruppene er imidlertid ikke statistisk signifikante ($t=-.621$, $df=75$, $p=.458$, $d=.024$), og nullhypoteses må derfor opprettholdes.

Tabell 10 – Mestringsforventninger til høyt-presterende elever født i ulike kvartal

	N	M	SD	MD	p	Cohens d
1.kvartal H	39	1.973	0.789			
				-0.021	.458	0.024
3.kvartal H	38	1.994	0.950			

Videre ble en tilsvarende hypotese for lavt-presterende elever formulert, også denne ensidig.

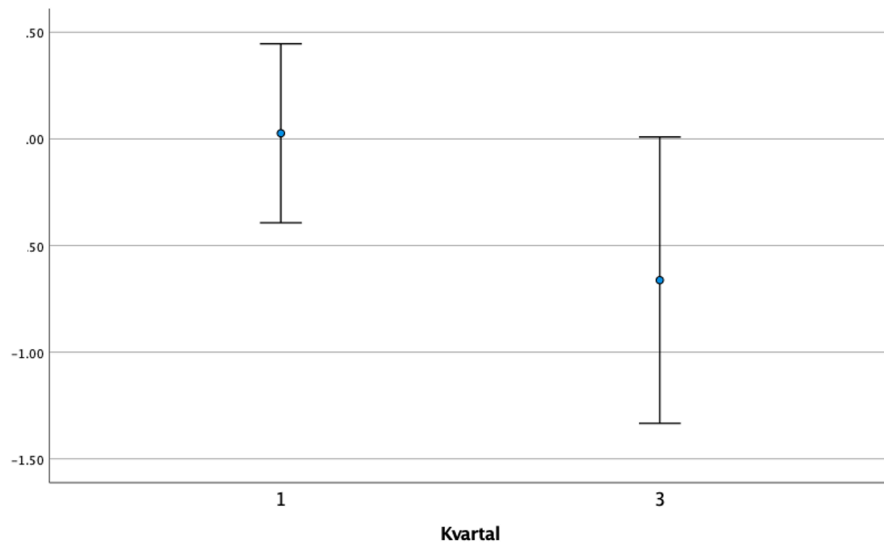
Eldre lavt-presterende elever har høyere mestringsforventninger i matematikk enn yngre lavt-presterende elever

Tabell 11 indikerer at lavt-presterende elever født i 1.kvartal (M=.027, SD=1.144) har signifikant høyere mestringsforventninger ($t=1.733$, $df=63$, $p=.088$) i matematikk sammenlignet med lavt-presterende elever født i 3.kvartal (M=-.662, SD=1.924). Forskjellen mellom gruppene er imidlertid kun statistisk signifikante på 0.1 nivå. På grunn av dette forkastes nullhypotesen, om at det ikke er forskjell i lavt-presterende elevers mestringsforventning i matematikk avhengig av hvilket kvartal de er født i. Cohens d utviser imidlertid en liten effektstørrelse, noe som kan tyde på at det er en svak, men reell forskjell mellom gruppene mestringsforventninger. Forskjellen mellom gruppene blir mer tydelig når man ser på spredningen til hver av gruppene. Spredningen i de yngre lavt-presterende elevenes rapporterte mestringsforventninger er vesentlig større sammenlignet med spredningen i de relativt eldre lavt-presterende elevens rapportering. Dette illustreres nedenfor i Figur 9.

Tabell 11 – Mestringsforventninger til lavt-presterende elever født i ulike kvartal

	N	M	SD	MD	p	Cohens d
1.kvartal L	31	.027	1.144			
				.0688	.088*	0.43
3.kvartal L	34	-.662	1.924			

(*:signifikant på 0.1 nivå, **:signifikant på 0.05 nivå)



Figur 8 - Gjennomsnitt og spredning mellom lavt-presterende elever født i 1. og 3.kvartal

5. Diskusjon

Hensikten med denne masterstudien var å undersøke om det var sammenheng mellom relativ alder og mestringsforventninger i matematikk hos elever på 9.trinn. For å kunne besvare problemstillingen ble den delt opp i flere forskningsspørsmål. Utviklingen av instrumentet, og valideringen av dette er viet et eget forskningsspørsmål. I tillegg til å undersøke alle elevene født i 1.kvartal og 3.kvartal, blir også undergrupper som kjønn og høyt- og lavt-presterende elever undersøkt med tanke på mestringsforventninger og RAE. I dette kapittelet blir resultatene diskutert i lys av teori og tidligere forskning, og presenteres systematisk etter de ulike forskningsspørsmålene. Siden jeg ikke har klart å finne tidligere forskning knyttet til mestringsforventninger i matematikk og RAE, vil funnene diskuteres opp mot teori og forskning knyttet til henholdsvis mestringsforventninger og prestasjon, og RAE og prestasjon.

5.1 Et instrument – mestringsforventninger i matematikk

Det første forskningsspørsmålet mitt var; *Hvordan vil et instrument som måler mestringsforventninger i matematikk hos elever på 9.trinn kunne se ut?*

I litteraturgjennomgangen fant vi ikke et instrument som var tilpasset for å måle mestringsforventninger i matematikk hos elever på dette trinnet i norsk kontekst.

Medlemmene i SESAM har derfor utviklet og validert et slikt instrument. Nedenfor vil jeg ta for meg hvilke grep som ble tatt i prosessen med å utvikle og validere instrumentet.

Pajares og Miller (1995) påpeker at det mest teoretisk passende og empirisk garanterte er at instrumentet som måler mestringsforventningene tar utgangspunkt i de samme ferdighetene som kreves for oppgavene som ofte løses i etterkant. Med andre ord er det viktig at det er samsvar mellom de ferdighetene som kreves for å løse oppgavene og de ferdighetene som det tas utgangspunkt i når man skal vurdere egne evner. I samsvar med dette, fulgte vi fremgangsmåten til Chen (2003) som baserer seg på det nevnte over. At undersøkelsens første- og siste del skulle basere seg på de samme items var viktig nettopp for å utvikle et teoretisk passende instrument, og dermed styrende for undersøkelsens design. Med inspirasjon fra Chen (2003), valgte vi å basere undersøkelsen på oppgaver fra TIMSS. TIMSS-undersøkelsen er designet slik at den skal fange opp bredden til faget slik det undervises i deltakerlandene, og baserer seg på informasjon om pensum og læreplaner knyttet til de ulike trinnene undersøkelsen gjennomføres på (IEA, u.d). Ved å ta i bruk allerede validerte matematikkoppgaver tilpasset elever på 9.trinn i norsk kontekst, hadde vi et solid

grunnlag som utgangspunkt for utviklingen av undersøkelsen. Dette kan regnes som en styrke ved undersøkelsen. Videre vil det også være viktig å nevne at alle medlemmene i SESAM har bidratt til å utvikle de tre instrumentene undersøkelsen besto av. Det å ha mulighet til å diskutere hvilke items en undersøkelse skal bestå av, ordlyden til de ulike items og selve oppbyggingen av undersøkelsen er viktig i prosessen med å utvikle spørreundersøkelser (Johannessen et al., 2012).

Et annet tiltak som ble gjennomført som en del av valideringen av instrumentet var pilot-studien. En pilot-studie har flere funksjoner, men først og fremst handler det om å bedre validiteten, reliabiliteten og de praktiske aspektene ved selve gjennomføringen (Cohen et al., 2018). Gjennomføringen og resultatene fra pilot-studien synliggjorde aspekter i undersøkelsen som krevde forbedringer. Blandt annet brukte elevene relativ lang tid på å løse oppgavene i undersøkelsen siste del. Analysene viste også at tre items ikke passet inn i det underliggende konstruktet (mestringsforventninger), og dermed ikke fulgte det samme mønsteret som de andre items. På bakgrunn av dette ble fire oppgaver fjernet, og den endelige undersøkelsen besto av tolv items. Gjennomføringen av piloten gjorde oss også bevisst hvor viktig tydelige instruksjoner er. Med andre ord gjorde vi oss verdifulle erfaringer basert på den praktiske gjennomføringen slik at denne kunne optimaliseres til hovedundersøkelsen.

Rasch-analyse ble brukt for å validere instrumentet. Ved å ta i bruk RSM i valideringsprosessen kunne vi sikre at instrumentet målte det det hadde til hensikt å måle. Rasch-analyser tillater en transformasjon av ordinale data, slik at data fra Liker-skalaer kan behandles mer presist enn ved klassisk testteori. Hvis vi hadde brukt de ordinale verdiene som grunnlag for videre analyser hadde man antatt at avstanden mellom svarkategoriene hadde vært lik, og at alle items er vektet likt, altså fremstår som like vanskelig. Ved å overføre dataene våre til Rasch-modellen gjør at vi kan transformerer dataene våre til et datasett som inneholder mye mer vektet informasjon, enn hvis vi hadde tatt utgangspunkt i klassisk testteori.

Basert på de ulike Rasch-analysene presentert i kapittel 4.1 er det grunnlag for å konkludere med at instrumentet som måler mestringsforventninger for elever på 9.trinn ikke inneholder en ekstra dimensjon. Instrumentet oppleves heller ikke annerledes for de ulike undergruppene i utvalget, og dermed er det tilstrekkelig invariant. På bakgrunn av dette kan man konkludere med at instrumentet måler det det var ment til, og dermed har god indre validitet. Analysene

knyttet til valideringsprosessen viste at konstruksjonen mellom det itemet som ble oppfattet som det vanskeligste og det itemet som ble oppfattet som enklest var godt dekket. Likevel viser materiale en skjevstilling (se Figur 5) hvor elevene med høyest mestringsforventninger er lokalisert over de items som oppfattes som vanskeligst. For en mer nøyaktig estimering av disse elevenes oppfattede kompetanse, kreves flere og nye items. Den deskriptive statistikken fra undersøkelsens mestringsdel viste også at ingen av respondentene fikk full skår på denne delen. I den sammenheng er det også viktig å nevne at det var god andel elever som fikk null poeng på denne delen av undersøkelsen. Om dette skyldes at mange elever ikke har prøvd å løse oppgavene, eller om oppgavene har vært for vanskelig er vanskelig å avgjøre. Det er likevel viktig å være bevisst at dette kan være en svakhet ved studien.

Selv om instrumentet oppfylder kravene for indre validitet, rettes det mer usikkerhet mot å avgjøre om dette også gjelder for ytre validitet. Ytre validitet omhandler resultatenes generaliserbarhet, og tar opp spørsmål angående for hvem resultatene er gyldige for, og for hvilke situasjoner de er gyldige i. Et grunnprinsipp for å kunne si at en undersøkelse har god ytre validitet er at utvalget er representativt for hele populasjonen. Siden denne undersøkelsen baserer seg på bekvemmelighetsutvalg og elevene som deltok ikke var tilfeldig trukket, er det ikke grunnlag for å si at utvalget er representativt for alle elever på 9.trinn i Norge. Videre undersøkelser som baserer seg på et større sannsynlighetsutvalg, er nødvendig for å kunne si at funnene fra denne undersøkelsen er representativt for hele populasjonen.

Valideringsprosessen var avgjørende for å kunne undersøke de øvrige forskningsspørsmålene. At instrumentet har gjennomgått en slik valideringsprosess gjør at andre kan ta det i bruk uten å være avhengig av å foreta en grundig pilotering og foranalyser. Det vil også være viktig å nevne at siden instrumentet er bygd opp av oppgaver som dekker de ulike emneområdene i TIMSS, gjør at instrumentet kan brukes i studier som ønsker å ha et fokus på disse, i større grad enn det jeg har i denne masteroppgaven. Et overblikk over innholdet i undersøkelsen i sin helhet viser at man kan få ut mange ulike typer data basert på de ulike instrumentene. Alle områdene er ikke brukt optimalt i denne masteroppgaven, men ved at vi for eksempel har inkludert de ulike emneområdene fra TIMSS viser at instrumentet har stort potensiale som videre kan brukes i andre studier.

5.2 Relativ alder og mestringsforventninger i matematikk

I denne delen av masteroppgaven diskuteres forskningsspørsmålet; *Hvordan er nivået på mestringsforventninger i matematikk sett i forhold til relativ alder?* For å diskutere dette vil jeg først ta for meg lærerens perspektiv, og se på hvilken rolle læreren kan ha med tanke på RAE og mestringsforventninger i matematikk. Videre vil jeg fokusere på elevperspektivet, hvor Banduras kilder til mestringsforventninger trekkes frem. Hva disse resultatene har å si for elevene, og når er egentlig det ideelle prøvetidspunktet, er også aspekt som trekkes frem.

Bakgrunnen for antagelsen om at relativt eldre elever har høyere mestringsforventninger i matematikk baserer seg på tidligere forskning som rapporterer at relativt eldre elever presterer bedre i matematikk (Aune et al., 2018; Bjerke et al., til fagfelleevaluering; Olsen & Bjørnsson, 2018) og at mestringsforventninger i stor grad predikerer elevens prestasjon i matematikk (Pajares & Miller, 1995; Randhawa et al., 1993). De deskriptive dataene fra undersøkelsen indikerer at de relativt eldre elevene har høyere mestringsforventninger i matematikk sammenlignet med de relativt yngre elevene. Resultatene fra analysene viser at denne forskjellen er statistisk signifikant ($p < 0.1$). Som nevnt i teorikapittelet er det vanlig å skille mulige årsaker til hvorfor RAE eksisterer mellom biologibaserte- og psykososiale hypoteser. Martin et al. (2004) trekker frem to biologibaserte hypoteser om hvorfor alderseffekt er et fenomen på tvers av skolesystemer. Hypoteser om prenatale variasjoners påvirkning på fosterets utvikling kan ikke brukes til å forklare den RAE som ses i min oppgave, da de yngste elevene er født i september, oktober, november og desember. Dermed kan ikke de biologibaserte hypotesene, som baserer seg på prenatale variasjoner brukes til å diskutere årsakene til RAE i Norge, og vil derfor ikke diskuteres videre i denne oppgaven. Nedenfor vil jeg derfor fokusere på årsaker til RAE innenfor det psykososiale området.

Mine resultater viste at det var forskjell mellom eldre og yngre elevers mestringsforventning i matematikk, og at denne forskjellen var statistisk signifikant. Likevel viser effektstørrelsen at sammenhengen mellom gruppene var svak. I praktisk betydning betyr signifikansverdien at forskjellen mellom gruppene ikke har oppstått ved tilfeldighet, men siden effektstørrelsen er svak kan man anta at variablene jeg har sjekket ikke kan legges til grunn for hele denne forskjellen vi ser. Det vil derfor være nødvendig med videre undersøkelser innenfor feltet, som i større grad kunne forklart hvorfor denne forskjellen har oppstått. Det er likevel viktig å notere seg at effektstørrelsen er svak, med tanke på mulige konklusjoner som tas på bakgrunn av dette resultatet. Nedenfor vil jeg se på ulike årsaker til hvorfor elever født tidlig på året

rapporterer om høyere mestringsforventninger i matematikk enn elever født sent. Jeg vil først se på hvilken rolle læreren kan ha, før jeg tar elevens perspektiv og ser på konsekvenser alderseffekten kan ha for elevene.

En mulig forklaring til at eldre elever har høyere mestringsforventninger, som mine resultater viser, kan være lærerens påvirkning på elevene. Tidligere nevnte jeg dette som selvpoppfyllende profeti, som kan være en faktor som bidrar til den varige RAE. Forskning gjort av Baker et al. (2015) viste at lærere oppfattet de relativt eldre elevene som mer faglig dyktig, og forventet at disse elevene skulle prestere bedre enn sine relativt yngre medelever. Det viste seg også at relativt yngre elever i større grad blir ansett som elever med lærevansker sammenlignet med de relativt eldre elevene. En slik merking vil kunne påvirke elevenes reelle oppfatning av egne evner (Campbell, 2013). Gledhill et al. (2002) påpeker også at om eleven, implisitt eller eksplisitt, blir informert om akademiske nivå, vil dette påvirke elevens oppførsel. Elevene godtar også ofte de rolle de blir tildelt, og utfører relativt høyere eller lavere nivå ut fra denne rollen. Det er også en betydelig større andel av de relativt yngre elevene som blir diagnostisert med ADHD, og elever med atferdsvansker blir ofte oppfattet som akademisk svakere og lærere forventer mindre av disse (Baker et al., 2015). Med andre ord kan lærerens oppfatning av elevene påvirke både elevens akademiske prestasjon og sosiale status. Hvis det er slik at lærere oppfatter de eldste elevene som mer faglig dyktig, og at yngre elever i større grad anses som elever med læringsvansker, er det klart at dette kan påvirke elevers mestringsforventninger. Disse årsakene kan være med å forklare noen av tendensen vi ser i denne undersøkelsen. Lærerens oppfatning av elevene er viktig for hvor godt de presterer faglig. Nedenfor tas dette videre opp ved å trekke frem Bandura (1977) sine kilder til mestringsforventninger, og hvordan disse kan påvirke elevene forskjellig avhengig av når de er født på året.

I følge Jussim et al. (2009) gir lærere mer emosjonell støtte, samt gir flere hint og faglig støtte til de elevene som forventes å prestere godt. Videre viser forskningen, som nevnt over, at lærere oppfatter de eldste elevene som mer akademisk dyktig og dermed forventer mer av disse elevene. Oppmuntring fra lærere eller andre en elev har tillit til kan øke elevens forventning om mestring. Dette er én av Banduras fire kilder til mestringsforventning. Hvis lærerne bruker mer tid med, gir mer faglig støtte og flere hint til de relativt eldre elevene, er det klart at dette er en kilde til økt mestringsforventning. Slik støtte vil også kunne føre til at disse elevene opplever suksess i større grad enn elevene som ikke får tilsvarende faglig- og

emosjonell støtte. Å lykkes i å løse en oppgave vil styrke elevenes mestringserfaring. Mestringserfaringer er den kilden som regnes til å ha mest påvirkning på mestringsforventninger. Ved å tilrettelegge for at elevene opplever mestring vil dette i stor grad kunne øke elevenes mestringsforventning. Jussim et al. (2009) fant også at lærere smilte mer, og var mer oppmuntrende mot elever som forventes å prestere godt. Dette er aspekter som også kan tenkes å påvirke elevenes psykologiske og fysiologiske tilstander, altså Banduras fjerde kilde til mestringsforventning. Denne kilden er knyttet til de fysiske og emosjonelle reaksjonene elevene opplever knyttet til spesifikke aktiviteter eller oppgaver. Elever som opplever stress og angst i forbindelse med matematikkfaget vil sannsynligvis tolke disse følelsene som et bevis på manglende kompetanse, og dermed synker mestringsforventningene for den aktiviteten. Om læreren viser mer støtte, både faglig og emosjonelt, til de relativt eldre elevene kan det være rimelig å anta at denne kilden vil påvirke deres forventning om mestring i mindre grad. Med andre ord kan det tolkes som at læreren kan være noe av årsaken til de forskjellene i mestringsforventninger som ses i denne masteroppgaven.

Hittil i dette delkapittelet har jeg tatt for meg hvilken rolle læreren kan ha for RAE, og hvordan Banduras fire kilder til mestringsforventninger kan påvirke elevene forskjellig avhengig av fødselsmåned. Nedenfor vil jeg se mer inngående på forskjellene mellom de eldste og yngste elevene. Til tross for at resultatene fra min undersøkelse viser at forskjellen i rapportert mestringsforventning i matematikk mellom elever født i 1.kvartal og elever født i 3.kvartal er statistisk signifikant, viser den statistiske styrken at sammenhengen er svak. Forskjellen i gruppenes rapporterte mestringsforventning i matematikk blir imidlertid mer tydelig når man ser på spredningen i elevenes resultater (se Figur 4). Det er vesentlig større spredning i de relativt yngre elevenes forventning om mestring sammenlignet med de relativt eldre elevene. Med andre ord er det større variasjon i de yngre elevenes rapporterte mestringsforventning i matematikk. Innenfor psykososial forskning er det foreslått at umodenhet ved skolestart kan være en av grunnene til variasjon i RAE (Campbell, 2013). Relativt yngre elever har mindre tid og erfaring fra førskole-institusjoner, og kan dermed ha en sosial, emosjonell og kognitiv ulempe ved skolestart. Hvis de relativt yngre elevene er mindre nevrologisk moden, kan dette være en mulig årsak til hvorfor de rapporterer om lavere mestringsforventninger og har større spredning i resultatene sine, slik mine resultater viser.

Å vurdere vanskelighetsgraden til en oppgave kan også være krevende. Matematikkoppgaver kan fremstå relativt vanskelige, men likevel ikke inneholde komplekse regneoperasjoner (Bandura & Schunk, 1981). Elever som er mindre moden kan være mindre rustet til å vurdere om de kan mestre den spesifikke oppgaven eller ikke. Dette kan også være en mulig årsak til at det er større spredning i de yngre elevenes resultater. Hvis det å være relativt eldre, og dermed mer moden ved skolestart, har positiv effekt både på selvtillit, ambisjoner og utvikling av sosiale ferdigheter (Crawford et al., 2013b), er det rimelig å tro at dette kan være en del av forklaringen til resultatene i denne masteroppgaven.

Som vi så ovenfor er umodenhet en mulig årsak til variasjoner for RAE. Jeg vil videre i dette, og i de to neste avsnittene, peke på andre mulige årsaker. Jeg vil diskutere når det ideelle prøvetidspunktet er, samt trekke frem hvilke konsekvenser det å være yngst i en klasse kan ha. Som nevnt over oppfattes relativt eldre elever flinkere akademisk, og lærere forventer mer av disse elevene. Dette til tross for at det er konkludert med at det ikke er forskjell mellom gjennomsnittlig IQ avhengig av når man er født på året (Gledhill et al., 2002). Videre ble det også konkludert med at lese- og skriveferdighetene til de yngste elevene var på det nivået man kunne forvente ut fra den alderen de var. Likevel viser undersøkelser gjort av Crawford et al. (2013b) at elever som er født sent på året har 5,4% større sannsynlighet for å motta en form for spesialundervisning, enn de relativt eldre elever. Elever som blir oppfattet som elever med læringsvansker, har i følge Gledhill et al. (2002) økt risiko for sekundære problemer som lav selvtillit, dårligere sosiale ferdigheter og mindre tro på egne evner. Slike sekundære problem kan også legges til grunn for at de yngre elevene rapporterer om lavere mestringsforventninger i matematikk enn de eldre elevene. Med andre ord kan det tolkes at det å være yngst i en klasse kan ha negative konsekvenser for elevenes prestasjon og mestringsforventning, men også andre sekundære problem som Gledhill et al. (2002) nevner.

En grunn til at de yngre elevene tildeles slike roller som nevnes over, kan være at de sammenlignes med kognitivt eldre og mer modne elever (NOU 2019: 3, 2019). Dette bekreftes også av Black et al. (2011) som fant at de eldste barna presterte bedre kun fordi de var eldre og mer modne på tidspunktet prøven ble gjennomført. En kan stille spørsmålet om når det ideelle prøvetidspunktet er. Så lenge prøven gjennomføres én gang, vil det jo alltid være en relativ aldersforskjell mellom elevene, som kan spille en avgjørende rolle for elevenes mestringsforventninger og prestasjon. Det kan derfor tenkes at de relativt yngre elevenes resultater fra denne undersøkelsen hadde vært annerledes, dersom elevene hadde

gjennomført undersøkelser på en tid hvor deres relative alder hadde vært lik som de eldste elevene. Hvilket resultat hadde jeg fått hvis elevene hadde gjennomført undersøkelsen i den måneden de fylte år? At januarbarn hadde gjennomført undersøkelsen i januar, mens novemberbarn hadde gjennomført den i november. En problemstilling man da vil komme over er å kontrollere om forskjellen i skår hadde skyltes relative alder, eller om de yngre elevene hadde lært mer siden de måtte testes på et senere tidspunkt og dermed har mer skoleerfaring. Dette er et spennende tankespinn, og noe som bør undersøkes videre. Nedenfor vil jeg ta for meg hvilke konsekvenser lave mestringsforventninger kan ha.

Forskning viser at mestringsforventninger har positiv effekt på akademisk engasjement, akademisk prestasjon, motivasjon og for valg av karriere (Bandura & Locke, 2003; Hackett & Betz, 1989; Zimmerman, 1995). Hvis det faktisk er slik at de eldste elevene har høyere mestringsforventninger, kan dette medføre at de relativt eldre elevene også er mer motivert, engasjert i faget, og i tillegg presterer bedre. Hackett og Betz (1989) fant at mestringsforventninger i matematikk hadde positiv effekt på karrierevalg, og at studenter som hadde høye mestringsforventninger i matematikk også i større grad valgte matematikkrelaterte utdanninger. Olsen og Bjørnsson (2018) skriver, i likhet med Lien et al. (2005) at RAE er størst ved skolestart, men at gapet mellom de relativt yngste og relativt eldste elevene fortsatt er betydelig ved utgangen av grunnskolen. Dette er også en tid hvor elevene tar avgjørende valg for videre- og høyere utdanning. Hvis det er slik, som mine resultater viser, at relativt yngre elever har lavere mestringsforventninger, kan dette ha store konsekvenser for valg om videre utdanning. Dette vises i forskning gjort av Solli (2017) som fant at relativt yngre elever hadde mindre sannsynlighet for å starte rett i videregående skole etter grunnskolen. I tillegg hadde de mindre sannsynlighet for å fullføre videregående skole på normert tid, samt å søke høyere utdanning sammenlignet med relativt eldre elever. Hun fant også at elevene som var født i januar hadde i gjennomsnitt 19% høyere karakterer enn elevene som var født i desember. Med andre ord kan det å være yngst i en klasse, og dermed ha lavere forventninger om mestring, være en betydelig ulempe. Dette gjelder både for valg til videregående skole, men også med tanke på høyere utdanning. Som nevnt over har mestringsforventninger en viktig rolle for hvilke valg vi tar, og hvis relativt yngre elever har lavere mestringsforventninger kan dette få negative konsekvenser for videre utdanning og senere yrkesliv. I den sammenheng er det viktig å påpeke at til tross for at det finnes store forskjeller i utdanning knyttet til når man er født i det akademiske året, finnes det lite bevis for at disse

effektene vedvarer i voksen alder (Black et al., 2011; Crawford et al., 2013a). Det er likevel et fenomen man bør være bevisst når man jobber i grunnskolen.

Enda forskjellen i rapportert mestringsforventning mellom relativt eldre og relativt yngre elever er statistisk signifikant, er effektstørrelsen som rapporteres mellom gruppene imidlertid lav. Dette kan tyde på at det er sammenheng mellom gruppene, men at denne sammenhengen er svak. Dette illustreres i Figur 5 hvor vi ser fordelingen av elevene og items, og elever født i 1.kvartal ikke er klart overrepresentert i det øverste siktet. Figur 5 viser også at de elevene som rapporterer om høyest mestringsforventninger i matematikk er elever født i 3.kvartal. En mulig forklaring til dette resultatet kan være at effekten av RAE og sammenhengen mellom mestringsforventninger og prestasjon varierer avhengig av alder. Street et al. (2017) fant at det var større sammenheng mellom prestasjon og mestringsforventninger for elever på 8. og 9.trinn enn for elever på 5.trinn. Dette underbygges også av Doménech-Betoret et al. (2017) som fant at mestringsforventninger hadde betydelig effekt på prestasjon for elever i 12-17 års alderen. Undersøkelser knyttet til RAE viser derimot at det er større sammenheng mellom relativ alder og prestasjon jo yngre elevene er. Olsen og Bjørnsson (2018) understreker dette, og skriver at RAE er to ganger større for elever på 4. og 5.trinn enn for elever på 8. og 9.trinn. Også Bjerke et al. (til fagfelleevaluering) fant en sammenheng mellom RAE og norske elevers prestasjon i matematikk i TIMSS 2015, men kun for elevene på 4., 5., og 8.trinn, og ikke for elever på 9.trinn. De fant også at forskjellen i matematikkskår mellom elever født tidlig og elever født sent var signifikant større på 4.trinn enn på 9.trinn. Basert på dette er det grunn til å anta at forskjellene fra min undersøkelse hadde vært enda mer signifikante, og sammenhengen sterkere om undersøkelsen hadde vært gjennomført på et lavere trinn. Det påpekes likevel av Olsen og Bjørnsson (2018) at estimatet for RAE er statistisk signifikante også blant de eldre elevene. Det hadde derfor vært interessant å undersøke dette videre, og sett om forskjellen fortsatt var statistisk signifikant ved utgangen av videregående skole. Det er i hvert fall viktig å merke seg muligheten for at sammenhengen mellom mestringsforventninger og RAE kunne vært annerledes om det ble undersøkt på et annet trinn, enn det som er gjort i denne studien.

5.3 Relativ alder og mestringsforventning knyttet til de fire emneområdene fra TIMSS

Dette delkapittelet vil fungere som et videre dypdykk for å undersøke om det er sammenheng mellom RAE og mestringsforventninger og har til hensikt å belyse følgende forskningsspørsmål; *Er det forskjeller i rapportert mestringsforventning til de ulike items sett i lys av emneområdene fra TIMSS, og relativ alder?* Forskjellene undersøkes ved hjelp av deskriptiv statistikk, altså kun en sammenligning av measure, item for item. Med andre ord er det ikke gjennomført gjennomsnittsanalyser for å undersøke om forskjellene mellom de ulike items er statistisk signifikante. Det er likevel spennende å undersøke om elevene har høyere mestringsforventninger til et emneområde fremfor et annet, og om vi ser forskjeller knyttet til når elevene er født på året.

Den delen av TIMSS-undersøkelsen som har til hensikt å kartlegge elevers kunnskap og forståelse er delt inn i de fire emneområdene tall, algebra, geometri og statistikk. Resultatene fra TIMSS 2015 viste at den norske profilen er preget av store forskjeller i skår mellom de ulike emneområdene (Bergem, 2016). I 2015 presterte de norske elevene godt i emneområdet tall, enda bedre innenfor statistikk, men betydelig lavere innenfor geometri (Bergem, 2016). Elevene skåret klart dårligst innenfor emneområdet algebra. Basert på disse resultatene og tidligere forskning om sammenhengen mellom mestringsforventninger og prestasjon, antok jeg at det også ville være forskjell i rapportert mestringsforventninger knyttet til de ulike item sett i lys av emneområdene. Elevene som var født i 1.kvartal rapporterte om høyere mestringsforventninger for seks items. Av disse tilhørte to items emneområdet tall, ett item tilhørte algebra, mens de resterende items tilhørte emneområdet geometri. Med andre ord rapporterte de relativt eldre elevene om høyere mestringsforventninger for alle items innenfor geometri. Elevene som var født i 3.kvartal rapporterte om høyere eller like høye mestringsforventninger for alle items innenfor emneområdet statistikk. Som nevnt over presterer norsk elever vesentlig dårligere innenfor emneområdet geometri enn innenfor statistikk. Et mulig tankespinn kan være at siden relativt eldre elever presterer bedre enn yngre elever, har de også høyere mestringsforventninger innenfor det emneområdet som er vanskeligst. Denne forklaringen kan ikke tas i bruk til å forklare hvorfor de yngre elevene har høyere mestringsforventninger for emneområdet statistikk. Om det er en årsak til at de relativt eldre elevene har høyere forventning om mestring innenfor geometri, men ikke innenfor statistikk kan ikke besvares basert på denne undersøkelsen, og videre dypdykk trengs for å

undersøke om denne tendensen er gyldig for hele populasjonen. Den deskriptive statistikken viser at det gjennomgående er små forskjeller i measure mellom eldre og yngre elever, med 0.17 logits som den største differansen, knyttet til hvert item innenfor de ulike emneområdene (se Tabell 5). Dette til tross for at det var store forskjeller i prestasjon mellom disse områdene både i TIMSS 2015 (Bergem, 2016) og TIMSS 2019 (Kaarstein et al., 2020). Selv om forskjellene synes relativt små, hadde det likevel vært fint å forfølge for å sjekke om forskjellene er statistisk signifikante.

I videre undersøkelser knyttet til dette forskningsspørsmålet ønsker jeg å se om tendensene fra prestasjon i TIMSS viser seg likt i denne studien. For å undersøke dette henvises det til Tabell 5 som viser rangeringen av opplevd vanskelighetsgrad for alle items som instrumentet besto av. Denne oversikten viser at item 1, innenfor emneområde tall, ble oppfattet som det vanskeligste itemet. Item 11, som også var fra emneområdet tall ble oppfattet som det letteste. Med andre ord inneholder emneområde tall både den oppfattet letteste og det oppfattede vanskeligste itemet. Som nevnt over har norske elever jevnt over prestert godt innenfor dette emneområdet tidligere, men elevene presterte signifikant dårligere innenfor emneområdet i 2019 enn i 2015. Det er derfor vanskelig å avgjøre om det er noen spesiell årsak til at item 1 oppfattes som vanskeligst, og item 11 oppfattes som lettest. Dette er et spennende funn, men forfølges ikke videre i min oppgave. Videre viser resultatene fra min studie at tre av de seks item som elevene hadde minst tiltro til å mestre var innenfor emneområdet algebra. Det kan dermed tolkes slik at algebra var det emneområdet elevene totalt sett hadde lavest mestringsforventninger til. Dette samsvarer med resultater fra Bergem (2016) som skriver at norske elever har en tydelig svakhet i matematisk kompetanse innenfor algebra.

5.4 Relativ alder og mestringsforventninger knyttet til kjønn

I dette delkapittelet vil jeg diskutere det fjerde forskningsspørsmålet; *Er det kjønnsforskjeller i rapportert mestringsforventning i matematikk og RAE. Eventuelt hvilke forskjeller er det?* Jeg vil gjennomgående knytte hovedfunnene til teorien nevnt i kapittel 2. På bakgrunn av en noe begrenset teorigjennomgang av mestringsforventninger og RAE knyttet til kjønn, vil jeg derfor også peke fremover på videre forskning som trengs for å forklare tendensene som vises i mitt utvalg.

Tidligere forskning om kjønnsforskjeller i matematikkfaget viser at gutter uttrykker større tro på egne matematiske evner, og har en mer positiv holdning til matematikk enn jenter (Pajares, 2005). Dette samsvarer med forskningen til Jensen og Nordtvedt (2013) som viste at gutter har høyere forventning om mestring, høyere selvoppfatning, større grad av utholdenhet og kapasitet i matematikk sammenlignet med jentene. Videre er forskningen angående RAE i stor grad konsistent i den retningen at alderseffekten er større for gutter (Olsen & Bjørnsson, 2018). Basert på dette forventet jeg også forskjeller mellom kjønnene for variabelen mestringsforventninger, hvor opplevelsen av egne evner står sentralt. Resultatene fra mine analyser viser derimot at dette ikke er tilfellet i denne undersøkelsen. Den deskriptive statistikken viser at relativt eldre jenter har svakt høyere mestringsforventninger enn guttene født i samme kvartal, altså motsatt av det man kunne forvente ut fra tidligere forskning. Denne forskjellen er imidlertid ikke signifikant, og effektstørrelsen som rapporteres er svært lav. Det er derfor ikke grunnlagt for å anta at årsaken til denne forskjellen utviser noe annet enn tilfeldigheter. Funnet samsvarer med resultatene fra Pajares og Kranzler (1995) og Chen (2003) som heller ikke fant noen kjønnsforskjell, hverken i prestasjon eller mestringsforventninger i matematikk.

Videre analyser viser imidlertid at de relativt yngre guttene rapporterer om høyere mestringsforventninger sammenlignet med de relativt yngre jentene. Denne forskjellen var vesentlig større enn mellom guttene og jentene født i 1.kvartal. Til tross for at forskjellen mellom de relativt yngre jentene og guttene ikke var statistisk signifikante, antyder effektstørrelsen likevel en svak effekt mellom gruppene, i guttenes favør. Dette funnet samsvarer i større grad med tidligere forskning hvor gutter i stor grad utviser høyere mestringsforventninger sammenlignet med jentene (Pajares, 2005; Randhawa et al., 1993). Som vist i resultatene fra PISA 2012 utviser gutter større tro på matematiske evner og er mer motiverte enn jentene. Dette kan være noe av årsaken til at de relativt yngre guttene også i denne undersøkelsen skårer høyere på mestringsforventninger. Dette diskuteres videre nedenfor.

Effektstørrelsen viser at det kan være en svak sammenheng mellom jenter og gutter født i 3.kvartal, og forskjellen blir enda mer tydelig når man ser på spredningen mellom gruppene (se Figur 6). De relativt yngre jentene har vesentlig større variasjon i tro på egne evner sammenlignet med guttene født i samme kvartal. Dette diskuteres også over, hvor relativt yngre elever har større variasjon i rapportert forventning om mestring sammenlignet med de

relativt eldre elevene. Som tidligere nevnt kan en mulig grunn til dette være at de relativt eldre elevene er mer moden, og dermed mer nøyaktig i sin vurdering av egne evner. Dette kan imidlertid ikke legges til grunn som årsak til at det er størres spredning i de relativt yngre jentenes resultater, sammenlignet med de relativt yngre guttene. En annen mulig årsak kan være, som vist i Aune et al. (2018) sine undersøkelser, at relativt yngre jenter er overrepresentert blant de lavt-presterende elevene, og at dette kan være noe av grunnen til at de relativt yngre jentene har lavere mestringsforventninger. Disse tendensene vises imidlertid heller ikke i denne undersøkelsen (se Tabell 8), da det var like mange lavt-presterende gutter som lavt-presterende jenter født i 3.kvartal. Tidligere forskning viser nemlig at det ikke er signifikante forskjeller i jenter og gutters prestasjon i matematikk (Nordtvedt, 2013). Dette samsvarer med mine resultater, som viser at det var tilnærmet like mange jenter og gutter som ble kategorisert som høyt-presterende og lavt-presterende innenfor hvert kvartal. Med andre ord er det vanskelig å avgjøre hvorfor de yngre jentene rapporterer om lavere mestringsforventninger i matematikk sammenlignet med de yngre guttene, mens forskjellene er betydelig mindre mellom elevene født i 1.kvartal. Videre undersøkelser trengs for å undersøke om disse tendensene vises også i et større sannsynlighetsutvalg, og hvilke årsaker som kan forårsake nettopp denne forskjellen.

Til nå i dette delkapittelet har jeg sett på forskjeller mellom kjønnene knyttet til RAE og mestringsforventninger i matematikk. I videre undersøkelser ble det også kjørt analyser både innad i kjønnene, altså mellom jenter født i ulike kvartal og gutter født i ulike kvartal, og mellom kjønnene født i ulike kvartaler. Gallagher og Kaufman (2005) påpeker at mye av forskningen innen matematikk og kjønn fokuserer på forskjellen mellom kjønnene, til tross for at forskjellene innad i kjønnene sannsynligvis er like store. På den ene siden viser forskning om RAE at alderseffekten gjennomgående er større for gutter enn for jenter (Solli, 2017). Basert på dette kunne man forventet at den største forskjellen skulle vært mellom relativt eldre jenter og relativt yngre gutter. På en annen side viser forskning om mestringsforventninger at gutter utviser høyere tro på egne evner, til tross for at det ikke påvises forskjeller i prestasjon mellom kjønnene (Pajares, 2005). Resultatene fra mine analyser viste derimot at den største forskjellen i rapportert mestringsforventning knyttet til kjønn, var mellom jenter født i 1.kvartal og jenter født i 3.kvartal. De relativt eldre jentene rapporterte om høyere mestringsforventninger enn de relativt eldre guttene, og de relativt yngre jentene rapporterte om lavere mestringsforventninger enn de relativt yngre guttene. Forskjellene mellom gruppene er imidlertid ikke statistisk signifikant. Til tross for at

resultatene av disse analysene viser at forskjellen mellom jentene ikke er statistisk signifikante, viser effektstørrelsen at det er en svak sammenheng mellom jenter født i 1.kvartal og jenter født i 3.kvartal. Dette resultatet bekrefter Gallagher og Kaufman (2005) sine antagelser om at forskjellene innad i kjønnene sannsynligvis er like store eller større enn mellom kjønnene. Denne sammenhengen ses ikke mellom kjønnene født i 1.kvartal, og heller ikke mellom guttene født i 1.kvartal og guttene født i 3.kvartal.

Resultatene fra min undersøkelse viste at det var større forskjeller mellom kjønnene for de elevene som var født i 3.kvartal enn mellom de elevene som var født i 1.kvartal. Det er viktig å stille spørsmål om hva dette egentlig betyr? Hvorfor har yngre gutter høyere mestringsforventninger enn yngre jenter, mens denne forskjellen ikke vises blandt de eldste elevene? Ville vi sette disse tendensene i et større utvalg, hvor elevene var tilfeldig trukket? Videre finner jeg også at den største forskjellen i rapportert mestringsforventninger er mellom de eldste og yngste jentene. Dette motsetter seg det man kunne forvente ut fra teori, hvor det mest logiske hadde vært at den største forskjellen var mellom gutter og jenter. Kan dette være tegn på at trenden om at gutter har høyere mestringsforventninger enn jenter snur? Basert på en noen begrenset litteraturgjennomgang knyttet til RAE og kjønn, og kjønn og mestringsforventninger har jeg over diskutert funnene opp mot den teorien som inkluderes i oppgaven. Likevel kan jeg ikke besvare spørsmålene som stilles i dette avsnittet. Hvis tendensene som kommer til syne i mitt utvalg, også stemmer før et større sannsynlighetsutvalg er dette noe som bør undersøkes nærmere i større studier.

5.5 Relativ alder og mestringsforventninger for høyt-og lavt-presterende elever

I denne delen av oppgaven fokuseres det på det siste forskningsspørsmålet; *Er det forskjell mellom høyt-presterende og lavt-presterende elevers mestringsforventning i matematikk knyttet til relativ alder? Eventuelt hvilke forskjeller er det?* Jeg vil i denne delen, i likhet med kapittel 5.2, diskutere funnene ut fra elevens perspektiv, hvilken rolle læreren kan ha og hvilke konsekvenser lave mestringsforventninger kan ha for elever.

Tidligere forskning viser at det er signifikant sammenheng mellom mestringsforventninger og prestasjon i matematikk (Bandura, 1997; Pajares & Kranzler, 1995). Mestringserfaringer

regnes som den kilden som har mest innvirkning på mestringsforventninger (Bandura, 1977). For elever som legger inn en innsats og opplever suksess, vil forventningene om å mestre en lignende aktivitet eller oppgave øke. Basert på teori og empiri antok jeg derfor å finne at de høyt-presterende elevene utviste høyere mestringsforventninger enn lavt-presterende elever. Mine resultater bekrefter tidligere forskning om at det er sammenheng mellom mestringsforventninger og prestasjon i matematikk. De høyt-presterende elevene rapporterte om betydelig høyere forventning om mestring sammenlignet med de lavt-presterende elevene, både innad i 1.kvartal og innad i 3.kvartal. Denne sammenhengen var statistisk signifikant ($p < .001$), og effektstørrelsen viser at det er sterk sammenheng. Dette diskuteres videre nedenfor.

Elever med lave mestringsforventninger trekker seg ofte unna oppgaver og situasjoner som krever mer kompetanse enn de opplever å være i besittelse av, mens elever med høy forventning om mestring lettere går løs på utfordringer og viser mer innsats og utholdenhet (Bandura, 1977, 1997; Zimmerman, 1995). Med andre ord er det ikke oppsiktsvekkende at det var signifikante forskjeller i forventning om mestring mellom høyt-presterende og lavt-presterende elever. De lavt-presterende elevene rapporterte om vesentlig lavere mestringsforventninger enn de høyt-presterende. Det er da vanskelig å ikke stille spørsmål om elevene som allerede har vurdert seg selv til å ikke mestre en oppgave, i det hele tatt gjør en innsats eller prøver å løse disse oppgavene i undersøkelsesens siste del. Som Bandura (1977) skriver; om man ikke tror man har evnene som skal til for å kunne oppnå ønsket resultat, er det heller ingen grunn til å prøve. Som tidligere nevnt ble elevene som fikk 1-2 poeng på undersøkelsesens siste del kategorisert som lavt-presterende. Med andre ord ble ikke elevene som fikk null poeng inkludert i denne gruppen. Denne avgjørelsen ble tatt på grunnlag av å unngå å ekskludere de elevene som ikke var interessert i å gjennomføre undersøkelsen. Dette var en vanskelig avgjørelse for SESAM å ta, da det mest sannsynlig er elever som burde bli inkludert i kategorien lavt-presterende som i denne undersøkelsen ble ekskludert.

Over ble spørsmål om de lavt-presterende elevenes innsats på undersøkelsesens del 3 stilt. Dette diskuteres også nedenfor, og videre i avsnittet vil jeg også se på hvilke konsekvenser lave mestringsforventningers kan ha. Collins (1982) og Bouffard-Bouchard (1990) skriver at dyktighet lett kan overstyres av tvil, i den grad at elever kan gjøre dårlig bruk av ferdighetene sine under omstendigheter hvor troen på seg selv undergraves. På den ene siden kan det være at de lavt-presterende elevene trekker seg unna og gjør mindre innsats for å løse oppgavene i

undersøkelsesens siste del, fordi de allerede har vurdert seg selv til å ikke mestre oppgavene. Elevene kan inneha ferdighetene som skal til for å løse den spesifikke oppgaven, men tvilen på seg fører til at de gjør dårlig bruk av ferdighetene sine. Elevene kan vite hvilken handling som fører til et bestemt utfall, men hvis en tviler på at en kan utføre denne handlingen, vil ikke denne informasjonen påvirke valgene elevene tar. Det er dette som kalles forventet utkomme (outcome expectations), som Bandura (1977) fremhever, sammen med mestringsforventninger, har stor betydning for valg en tar. På den andre siden kan de lavt-presterende elevene rapportere om lave mestringsforventninger basert på at de faktisk mangler ferdighetene som skal til for å løse oppgavene. For disse elevene vil ikke forventningen om mestring kunne påvirket resultatet i stor grad, da de ikke har forutsetningene for å kunne mestre oppgavene i utgangspunktet. I den sammenheng er det viktig å bemerke at selv om de lavt-presterende elevene rapporterer om lave forventninger om mestring i matematikk, betyr ikke det at disse elevene har lav selvtillit. I følge Bandura (1977) forveksles ofte konseptene mestringsforventning og selvtillit. Selv om elevene vurderer seg selv til å ikke ha de ferdighetene som trengs for å løse oppgavene, rapporterer de ikke hva de føler om seg selv på generell basis (Zimmerman, 1995). Nedenfor diskuteres dette ytterligere, og variasjonen mellom høyt-presterende og lavt-presterende elevers mestringsforventning tas opp.

Forskjellen mellom høyt-presterende og lavt-presterende elever blir enda mer tydelig når man ser på spredningen mellom gruppene. Figur 7 viser at spredningen i resultatene er vesentlig lavere for høyt-presterende elever enn for lavt-presterende elever. Med andre ord er det større spredning i de lavt-presterende elevenes rapportering av mestringsforventninger, enn det er hos det høyt-presterende elevene. Dette bekrefter funnene til Chen (2003) og Pajares og Graham (1999) om at høyt-presterende elever er mer nøyaktig i sin vurdering av egne evner. De fant også at det var flere lavt-presterende elever som overvurderer egne evner, og dette kan være en mulig årsak til at det er større spredning blant denne elevgruppen. I følge Bandura og Schunk (1981) er det ingen enkel sak i å vurdere om man er i stand til å løse en oppgave eller ikke. Mange matematikkoppgaver kan fremstå som relativt enkle, men likevel inneholde komplekse regneoperasjoner. I tillegg skal denne vurderingen skje innenfor en svært begrenset tidsperiode, hvor elevene kun får se den aktuelle oppgaven i et fåtall sekunder. Mange elever er derfor ikke klar over eller i stand til å avgjøre hvilke ferdigheter som behøves for å løse en spesifikk oppgave. I tilfeller hvor det er uoverensstemmelser mellom forventet mestring og faktisk mestring, kan dette like ofte stamme fra misforståelser av oppgavens

vanskelighetsgrad, så vel som dårlig selvinnsikt (Bandura & Schunk, 1981). Dette kan være en av grunnene til at høyt-presterende elever er mer nøyaktig i sine vurderinger. De har bedre faglige forutsetninger for å vurdere hvilke ferdigheter og kunnskap som kreves for å løse en spesifikk oppgave sammenlignet med de lavt-presterende elevene. Dette kommer også frem underveis i undersøkelsen hvor en elev kom bort til meg og sa at han trodde oppgavene var lettere enn de viste seg å være. Han lurte derfor på om han kunne endre svarene sine i undersøkelsen første del, da han mente det ikke stemte lengre. Dette samsvarer med Bandura og Schunk (1981) forslag om hvor krevende det egentlig kan være å vurdere vanskelighetsgraden til en spesifikk matematikkoppgave. En annen mulig årsak til at det ikke er fullstendig samsvar mellom forventet mestring og faktisk mestring er det Kleven og Hjørdemaal (2018) betegner som observatøreffekten. I undersøkelser som innebærer at respondentene skal vurdere eller fortelle om seg selv, vil det alltid være en risiko for å fremstille seg selv mer positivt. Man kan da fremstille seg selv slik man ønsker man var, fremfor hvordan man egentlig er. Denne effekten vil variere avhengig av hvilken situasjon man er i, men vil alltid være en faktor som en må tenke over.

Som vi så ovenfor viser mine resultater at de høyt-presterende elevene har høyere mestringsforventninger i matematikk enn de lavt-presterende elevene, både for elevene født i 1.kvartal og 3.kvartal. Videre viser forskning innenfor RAE at relativt eldre elever presterte bedre enn relativt yngre elever. Det var derfor interessant å undersøke om vi så disse tendensene når man kun så på høyt-presterende elever eller lavt-presterende elever, født i ulike kvartal. Mine resultatet viste at det ikke var noen signifikant forskjell i mestringsforventninger mellom høyt-presterende elever født i 1.- og 3.kvartal. Med andre ord er det ikke funnet en RAE mellom høyt-presterende elever på 9.trinn når det kommer til mestringsforventninger i matematikk. Mellom lavt-presterende elever derimot var det en signifikant forskjell ($p < .1$), i den retningen at relativt eldre elever rapporterer om høyere mestringsforventninger enn de relativt yngre elevene. Dette til tross for at begge gruppene hadde samme skår på mestringsdelen av undersøkelsen (1-2 oppgaver riktig). En mulig årsak til dette kan være, som Baker et al. (2015) foreslår, at lærere anser relativt eldre elever som mer faglig dyktig sammenlignet med sine yngre medelever. Lærers oppfatning av elevene kan påvirke elevene, og deres kilder til mestringsforventninger. Som vi så ovenfor kan lærere ha innflytelse på kildene mestringserfaring, verbal overtalelse og psykologisk og fysiologisk tilstander. Hvis de relativt eldre elevene har opplevd mer faglig- og emosjonell støtte fra lærerne sine, kan dette påvirke elevenes forventning om mestring. Dette kan legges til grunn

som en mulig forklaring til hvorfor relativt eldre elever har høyere mestringsforventninger enn relativt yngre elever. Ut fra disse resultatene er det rimelig å anta at RAE er mer fremtredenene blant lavt-presterende elever, og at gode ferdigheter i matematikk kan fungere som en beskyttende faktor for RAE når det kommer til mestringsforventninger i matematikk. Videre undersøkelser knyttet til dette trengs for å kunne slå fast om konsekvensene av RAE er gjennomgående større for lavt-presterende elever enn for høyt-presterende elever.

At læreren kan ha en viktig rolle i at RAE er et fenomen vises gjennom hele oppgave. I uformell samtale med en lærer underveis i gjennomføringen av undersøkelsen på en skole ga han uttrykk for at han ikke var overrasket over hvilke elever som fortsatt jobbet med oppgaven i undersøkelsens siste del, og signaliserte at dette var elever som var flinke i matematikkfaget. Ut fra dette kan man tolke at de elevene som jobbet lenge og viste innsats, var elever som også hadde gode mestrings erfaringer i matematikkfaget fra før. Som Bandura (1977) skriver er mestrings erfaringer den kilden som har størst innvirkning på mestringsforventninger. Elever som allerede blir regnet som flinke har mest sannsynlig flere positive opplevelser med matematikkfaget, noe som kan føre til at de ikke gir seg like lett i møte med utfordringer. I tillegg til at høyt-presterende elever i større grad mestere faget, viser undersøkelser gjort av Jussim et al. (2009) at lærere viser mer støtte, både faglig og emosjonelt, til elever som forventes å prestere godt. Det viste seg også at lærerne gir flere hint når disse elevene møtte på utfordringer, smiler mer og var mer oppmuntrende til høyt-presterende elever. Hvis det faktisk er slik at elever som forventes å prestere godt får mer oppmerksomhet og tid med læreren kan dette være med å forklare resultatet vi ser i denne undersøkelsen. Med andre ord kan læreren spille en avgjørende rolle for elevenes utvikling av mestringsforventninger innenfor spesifikke aktiviteter, men også spille en viktig rolle for hvor store konsekvenser RAE kan få.

I likhet med undersøkelsene av spredningen blandt de de andre undergruppene, har også yngre lavt-presterende elever vesentlig større spredning i sine resultater sammenlignet med de relativt eldre lavt-presterende elevene (se Figur 8). Dette resultatet er gjennomgående for hele undersøkelsen, i den retningen at elever som er født i 3.kvartal har større variasjon i vurderingen av egne evner enn elever som er født i 1.kvartal. Med andre ord er de relativt eldre elevenes resultater mer samlet rundt gjennomsnittet enn de relativt yngre elevene. Det er flere mulig årsaker som kan legges til grunn for dette resultatet. Mange av disse diskuteres ovenfor, men jeg vil også kort nevne de her. Som Campbell (2013) påpeker kan umodenhet

være en årsak til variasjonene vi ser i RAE. De yngste elevene kan være opp mot 12 måneder yngre enn de eldste i årskullet. De relativt yngre elevene har i de fleste tilfeller mindre førskole-erfaring, noe som kan føre til at de er mindre rustet til de krav som skolen stiller. Crawford et al. (2013b) påpeker også at det å være eldst i en klasse kan ha positiv effekt både på selvtillit og ambisjoner, men også at de eldste elevene er mer nevrologisk moden. Dette kan være med å forklare tendensene vi ser i denne undersøkelsen, nemlig at de yngre elevene har lavere mestringsforventninger og at de har vesentlig større spredning i tro på egne evner. En annen mulig årsak som kan legges til grunn er vurdering av en oppgavens vanskelighetsgrad, som er helt sentral i studier som undersøker mestringsforventninger. Som nevnt ovenfor er det krevende å vurdere vanskelighetsgrad til en spesifikk matematikkoppgave, og i tilfeller der det ikke er samsvar mellom rapportert mestringsforventninger og faktisk mestring like ofte kan stamme fra misforståelser om vanskelighetsgrad, så vel som overvurdering av egne evner. Hvis yngre elever er mindre moden, kan dette også være en ulempe ved vurdering av oppgavers vanskelighetsgrad.

5.6 Praktiske implikasjoner

Det er gjennomført en god del forskning om RAE i skolen. Disse resultatene samsvarer i stor grad, i den retningen at relativt yngre elever presterer dårligere sammenlignet med de relativt eldre elevene. Videre viser forskningen at mestringsforventninger har positiv effekt på akademisk engasjement og akademisk prestasjon. Resultatene fra denne undersøkelsen viser at det også er sammenheng mellom RAE og mestringsforventninger i matematikk hos elever på 9.trinn, i den retningen at relativt eldre elever har høyere mestringsforventninger enn relativt yngre elever. Det er omstridt om hvilke årsaker som kan legges til grunn for RAE, men forskning gjort av Baker et al. (2015) viser at lærere oppfatter relativt eldre elever som mer faglig dyktig, og at lærere bruker mer tid og gir mer oppmerksomhet til disse elevene (Jussim et al., 2009). Med tanke på disse resultatene kan en mulig tolkning være at dersom lærere blir bevisst på dette, kan undervisningen legges til rette slik at alle elever får mer støtte. Som Bandura (1977) trekker frem er mestringserfaringer den faktoren som har størst innvirkning på mestringsforventninger. Dermed bør et mål for all undervisning være at elevene skal oppleve suksess og mestring. Læreren kan gjøre dette ved å tilrettelegge undervisningen, og gi elevene oppgaver som er autentiske og tilpasset deres nivå. Læreren må også være bevisst de andre kildene, og kanskje spesielt kilden verbal overtalelse. Ved å gjøre lærere bevisst på at RAE er et fenomen som påvirker både prestasjon og motivasjon

(mestringsforventninger) kan man i større grad unngå at lærerens påvirkning kan forsterke konsekvensene av RAE.

Som vi så ovenfor viser mine resultater at relativt eldre elever har høyere forventning om mestring sammenlignet med de relativt yngre elevene. Det ble imidlertid ikke funnet noen forskjell i mestringsforventninger mellom høyt-presterende elever født i ulike kvartal. Med andre ord var det ikke en RAE mellom relativt eldre høyt-presterende elever og relativt yngre høyt-presterende elever i min undersøkelse. Dette kan tolkes som at gode faglige ferdigheter kan beskytte elevene mot de negative konsekvensene av RAE. Hvis det å fokusere på å utvikle gode ferdigheter kan føre til at RAE får mindre konsekvenser, er dette et funn som kan være viktig for lærere å ha kjennskap til.

Gjennomgående der det er signifikante forhold, er effektstørrelsen lav for de fleste kombinasjoner, sett bort fra sammenhengen mellom lavt-presterende og høyt-presterende elever. Effektstørrelsen forteller om hvor stor del av den variasjonen som kan forklares av mine variabler. Selv om sammenhengen som rapporteres er svak, er det en sammenheng som det er verdt å vite om. Konseptet om RAE og konsekvensene av dette er lite belyst, og jeg fikk øynene opp for dette først etter over fire år på lærerutdanningen. At RAE er en faktor som kan påvirke elevens prestasjon og motivasjon er noe alle lærere bør vite. RAE er noe som burde være diskutert i lærerutdanningen.

5.7 Metodiske betraktninger

Som all annen forskning, har også denne studien sine begrensninger. Jeg vil i denne delen peke på disse begrensningene, men også trekke frem de aspektene som kan regnes som en styrke i studien.

Ved bruk av en kvantitativ spørreundersøkelse basert på et relativt stort utvalgt kan man med god sikkerhet konstaterte at det er forskjell mellom gruppene, men årsaksforholdene er det vanskelig å si noe om. Utvalget for denne studien baserer seg på bekvemmelighetsutvalg. En slik sluppmessig utvelgning kan føre til at noen grupper kan bli overrepresentert. I metodelitteraturen vektlegges det at man bør være forsiktig med å generalisere ut fra såkalte ikke-sannsynlighetsutvalg (Ringdal, 2013). Det bør derfor tas høyde for at resultatene fra denne masteroppgaven ikke kan generaliseres til å gjelde alle elever på 9.trinn i Norge.

Utvalget min undersøkelse baserer seg på består imidlertid av respondenter fra både bygd og by, og representerer fire ulike fylker. Det vil også være viktig å påpeke at hverken geografisk beliggenhet eller rektors prioriteringer påvirker elevenes motivasjon for å delta i undersøkelsen. Alle elevene som deltok, fikk informasjon om at det var frivillig å delta. På bakgrunn av dette kan man likevel regne med utvalget til en viss grad er representativt for populasjonen.

Det bør også nevnes at til tross for at undergruppene som besto av elever født i 1.kvartal og 3.kvartal besto av rundt 150 respondenter, besto undergrupper av disse igjen av et svært begrenset antall. Antall respondenter som ble regnet som høyt-presterende og lavt-presterende elever født i de ulike kvartalene var i overkant av 30, noe som regnes som et absolutt minimum (Johannessen et al., 2012). For å kunne trekke endelige konklusjoner basert på sammenhengen mellom høyt-og lavt-presterende elever, mestringsforventninger og RAE trengs et større utvalg, enn det som forekommer i denne undersøkelsen. Undersøkelsen kan likevel belyse mulige tendenser som vises i utvalget.

Man kan heller ikke være sikre på at elevene oppfatter spørsmålene på samme måte, og det må derfor tas høyde for at dette kan variere. Ved bruk av Likert-skalaer i undersøkelser, kan man ikke garantere at elevene tolker svarkategoriene likt. En elev kan bruke ett sett med kriterier for å velge svaralternativ fire, mens en annen elev kan bruke akkurat de samme kriteriene for å velge alternativ fem. Dette er som nevnt tidligere en av svakhetene ved bruk av Likert-skalaer som svarkategorier i undersøkelser. Ved at gruppemedlemmene i SESAM gjennomførte undersøkelsen på alle skolene, fikk elevene mulighet til å henvende seg direkte til en av oss dersom noe var uklart i undersøkelsen. På denne måten kunne vi sikre at elevene hadde tilgang til tilnærmet lik informasjon, og dermed senke sannsynligheten for systematiske målefeil. I undersøkelser der respondentene skal vurderer seg selv eller fremstille seg selv er det imidlertid mer vanlig med systematiske målefeil, noe som kan skape validitetsproblemer (Ringdal, 2013). I slike undersøkelser kan respondentene beskrive seg selv i mer positiv retning, og dermed vurdere seg selv slik de ønsker å være eller slik de oppfatter det er sosialt ønskelig, fremfor hvordan de egentlig er. At det ikke er fullstendig samsvar mellom respondentenes rapporterte mestringsforventning og faktisk mestring er vanlig i studier der mestringsforventninger blir undersøkt. Dette kan i følge Bandura og Schunk (1981) være på grunn av at det er vanskelig å avgjøre hvor kompleks en matematikkoppgave er på relativt kort tid, så vel som dårlig selvinnsikt. Ved at en av medlemmene i SESAM var til stede under

gjennomføringen fikk vi mulighet til å informere og forsikre elevene om at undersøkelsen var helt anonym, og at læreren ikke ville få tilgang til resultatene. På bakgrunn av dette kan det tenkes at denne feilkilden i liten grad truer validiteten til denne undersøkelsen.

At jeg har benyttet meg av Rasch-analyse for å validere instrumentet som måler mestringsforventninger, gjør at jeg med sikkerhet kan si at undersøkelsen faktisk belyser det den har til hensikt å gjøre. Som nevnt over kan man ikke anta at intervallene mellom kategoriene i en Likert-skala er like. Dette betyr at man ikke kan ta for gitt at følelsesintensiteten mellom *veldig enig* og *enig*, på en eller annen måte samsvarer med *veldig uenig* og *uenig*. En måte å motvirke denne begrensningen ved bruk av Likert-skalaer, er å ta i bruk moderne testteori. Moderne testteori behandler slike skaleringsproblemer mer presist enn klassisk testteori, og gir oss intervaller mellom kategoriene uten antagelser om at kategoriene er av samme størrelse. Det kan dermed regnes som en styrke å ta i bruk moderne testteori i valideringen av måleinstrumentet.

Til tross for at instrumentet som måler mestringsforventninger er validert, vil det være viktig å påpeke at det alltid kan optimaliseres. Som nevnt under 4.1 viser materialet en skjevstilling, i den grad at elevene med høyest rapportert mestringsforventning er lokalisert over de items som oppfattes som vanskeligst (se Figur 5). Dette betyr at nye items kreves for å kunne måle denne gruppens mestringsforventning i matematikk mer nøyaktig, enn den er målt i denne undersøkelsen. Det er også verdt å nevne at det ikke er gjort analyser for å validere undersøkelsens siste del, altså måling av elevens prestasjon. På bakgrunn av at denne delen besto av matematikkoppgaver fra TIMSS var det analyser av mestringsforventningsinstrumentet som var styrende fra pilot-studien til hovedundersøkelsen. Tolv oppgaver fra TIMSS utgjorde undersøkelsen første og siste del. TIMSS-undersøkelsen består av langt flere oppgaver som grunnlag for å kunne måle norske elevers kompetanse i matematikk. Med andre ord kan man ikke si at vår undersøkelse måler elevens kompetanse innenfor matematikk, slik som den måles i TIMSS. Siden denne delen av undersøkelsen var utgangspunktet for å kategoriseringen av høyt-og lavt-presterende elever, er dette nettopp viktig å være klar over. I denne kategoriseringen ble elevene som fikk 1-2 poeng regnet som lavt-presterende. Med andre ord ble ikke elevene som fikk null poeng inkludert i denne gruppen. Dette ble gjort for å unngå å inkludere de elevene som ikke prøvde å løse oppgavene i det hele tatt. Det er likevel grunn til å tro at noen av disse elevene burde vært inkludert i denne kategorien. Dette kan regnes som en svakhet ved datamaterialet. Det var heller ingen

elever som fikk full skår på mestringsdelen. Man kan derfor stille spørsmål om oppgavene var for vanskelig, til tross for at de var tilpasset 9.trinn i Norge.

6. Avslutning

Både mestringsforventninger og RAE er forskningsfelt som har fått mye oppmerksomhet, både nasjonalt og internasjonalt. Forskingen viser at både mestringsforventninger og RAE har innvirkning på elevenes prestasjon i matematikk. Sammenfattingen av disse to forskningsfeltene har, etter det jeg har funnet, imidlertid fått lite oppmerksomhet.

Utgangspunktet for denne masteroppgaven var et ønske om å sammenfatte disse to forskningsområder, for å undersøke om RAE også påvirker elevens motivasjon i matematikk. Dette resulterte i en kvantitativ studie av elevers mestringsforventning i matematikk, hvor jeg har sammenlignet elever født i 1.kvartal og 3.kvartal. Masteroppgaven har

Siden vi i litteraturen ikke fant et instrument tilpasset elever på 9.trinn i norsk kontekst har vi, gjennom forskningsgruppa SESAM, utviklet et slik instrument. Dette prosessen har blitt viet et eget forskningsspørsmål og fått en stor plass i oppgaven. Å utvikle og validere et instrument er et arbeid vi ikke har tatt lett på, og denne prosessen har gitt meg innsikt i hvor vanskelig og tidkrevende det er å lage et godt måleinstrument. Å være delaktig i å utvikle instrumentet som mine analyser baserer seg på har vært spennende, og det er gøy å se at det har fungert til sin hensikt. Ved å ta i bruk instrumentet fikk jeg mulighet til å besvare de øvrige forskningsspørsmålene som oppgaven baserer seg på. Dette har resultert i at jeg har funnet at det er signifikant sammenheng mellom når elever er født og hvor høy forventning om mestring de har i matematikk, i den retningen at relativt eldre elever rapporterer om høyere mestringsforventninger. Det er imidlertid viktig å trekke frem at selv om forskjellen er signifikant, viser effektstørrelsen at sammenhengen er svak. Med andre ord er forskjellen i mestringsforventninger i matematikk mellom elever født i 1.kvartal og elever født i 3.kvartal reell, men det er vanskelig å avgjøre hvor stor betydning den har i praksis

Videre viste forskningsresultatene at det ikke var signifikante kjønnsforskjeller, hverken mellom eller innad i kjønnene. Effektstørrelsen til analysene viste imidlertid at det er svak sammenheng mellom jenter og gutter født i 3.kvartal. Med andre ord er det tendenser til at yngre gutter har høyere forventning om mestring enn yngre jenter, mens denne forskjellen ikke vises blandt de eldre elevene. Dette var et spennende funn, og videre undersøkelser trengs for å bekrefte om dette også gjelder for hele populasjonen. Funnene fra denne studien støtter også Banduras (1997) teori om at det er sammenheng mellom mestringsforventninger og prestasjon innenfor matematikk. De høyt-presterende elevene rapporterte om signifikant

høyere forventning om mestring, sammenlignet med de lavt-presterende elevene, både innenfor 1.kvartal og 3.kvartal. Kanskje det mest oppsiktsvekkende funnet er at jeg har identifisert en RAE mellom lavt-presterende elevers mestringsforventning i matematikk, i den retningen at eldre lavt-presterende elever har høyere mestringsforventninger enn yngre lavt-presterende elever. Denne forskjellen ble ikke funnet mellom høyt-presterende elever. Man kan derfor stille seg spørsmålet om RAE er mer utbredt blant elever som presterer svakt, enn elever som er faglig sterke. Eller om gode faglige ferdighet kan virke som en beskyttende faktor for RAE.

Hovedfunnene fra analysene har blitt diskutert opp mot teori og tidligere forskning. Jeg har også, der det har latt seg gjøre, pekt på funn som trenger ytterligere forskning for å kunne forklares. Gjennomgående har Banduras (1977) teori om mestringsforventninger, og de fire kildene til mestringsforventninger vært sentrale for å diskutere mulige årsaker til de tendensene vi ser i oppgaven. Jeg har også belyst hvilken rolle læreren har, og hvordan vi som lærere kan være med å forsterke RAE ved å behandle elevene forskjellig. Kunnskap om at RAE er et fenomen på tvers av skolesystemer er noe jeg tar med meg inn i lærerhverdagen. Å få innsikt i de konsekvensene det å være yngst i en alderskohort kan ha, er viktig å være klar over som lærer. Jeg sitter igjen tanker om at RAE er noe som i større grad må belyses, både i lærerutdanningen, men også for alle som jobber i norsk grunnskole. At lærere kan ha en viktig rolle i forebyggingen av de negative konsekvensene RAE har, som lave mestringsforventninger, er jeg overbevist om. Det kreves likevel mer forskning på området for å endelig etablere hvilke årsaker som kan legges til grunn for RAE.

6.2 Videre forskning

Denne masterstudien har bidratt med et nytt instrument for å undersøke norske elevers mestringsforventninger i matematikk. Den har også bidratt med å kaste lyst over at det faktisk kan være en sammenheng mellom elevens fødselsmåned og hvor høye mestringsforventninger de har i matematikk. Funnene mine underbygger også behovet for mer forskning på dette feltet. I fremtidige undersøkelser hadde det vært interessant å se om andre studier får resultatene som er sammenfallende med mine. Til tross for at sammenhengen mellom elever født i 1.kvartal og 3.kvartal var statistisk signifikant, viser effektstørrelsen at sammenhengen mellom gruppene var svak. Det er derfor nødvendig å undersøke om det er andre faktorer som i større grad forklarer sammenhengen mellom gruppene. Det kunne også vært interessant å gå

dypere inn i RAE blant lavt-presterende elever, for å undersøke om sammenhengen mellom RAE og mestringsforventninger gjennomgående er større for disse elevene, og om gode faglige ferdigheter i noen grad motvirker konsekvensene av RAE. Intervju- og observasjonsstudier kunne vært aktuelle metoder å benytte, for å kunne få dypere innsikt i mulige årsaker til sammenhengen mellom RAE og mestringsforventninger. Det hadde også vært nyttig å få gjennomført undersøkelsen på et annet klassetrinn, for å undersøke om sammenhengen hadde vist seg annerledes ved en annen alder enn den som undersøker i denne oppgaven.

Litteraturliste

- Aune, T. K., Ingvaldsen, R. P., Vestheim, O. P., Bjerkeset, O. & Dalen, T. (2018). Relative age effects and gender differences in the national test of numeracy: a population study of Norwegian children. *Frontiers in Psychology, 9*:1091.
- Aune, T. K., Pedersen, A. V., Ingvaldsen, R. P. & Dalen, T. (2017). Relative Age Effect and Gender Differences in Physical Education Attainment in Norwegian Schoolchildren. *Scandinavian Journal of Education Research, 61*(3), 369-375.
- Backe-Hansen, E., Walhovd, K. B. & Huang, L. (2014). *Kjønnsforskjeller i skoleprestasjoner - En kunnskapsoppsummering (5/2014)*. Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.
- Baker, C. N., Kupersmidt, J. B., Tichovolsky, M. H., Voegler-Lee, M. E. & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of educational psychology, 107*, 805-820.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.
- Bandura, A. (1989). Human agency in social cognitiv theory. *American Psychologist, 44*, 1175-1184.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for creating self-efficacy scales. I F. Pajares & T. Urdan (Red.), *Self-efficacy beliefs of adolescents*. Greenwich, Connecticut: IAP.
- Bandura, A. & Locke, E. A. (2003). Negativ self-efficacy and goal effects revisited. *Journal of applied psychology, 88*, 87-99.
- Bandura, A. & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest throug proximal self-motivation. *Journal of Personality and Social Psychology, 41*, 586-598.
- Bergem, O. K. (2016). Hovedresultater i matematikk. I O. K. Bergem, H. Kaarstein & T. Nilsen (Red.), *Vi kan lykkes i realfag - Resultater og analyser fra TIMSS 2015*. Universitetsforlaget.
- Bergem, O. K., Kaarstein, H. & Nilsen, T. (2016). TIMSS 2015. I O. K. Bergem, H. Kaarstein & T. Nilsen (Red.), *Vi kan lykkes i realfag - Resultater og analyser fra TIMSS 2015*. Oslo: Universitetsforlaget.
- Bjerke, A. H. & Eriksen, E. (2016). Measuring pre-service teachers' self-efficacy in tutoring children in primary mathematics: an instrument. *Research in mathematics education, 18*, 61-79. <https://doi.org/10.1080/14794802.2016.1141312>
- Bjerke, A. H., Smestad, B., Eriksen, E. & Rognes, A. (til fagfelle vurdering). Relationship Between Birth Month and Mathematics Performance in Norway.
- Black, S. E., Devereux, P. J. & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The review of economics and statistics, 93*, 455-467.
- Bond, T. G. & Fox, C. M. (2015). *Applying the rasch model: Fundamental Measurement in the Human Sciences* (Tredje. utg.). New York and London: Routledge.
- Bong, M. & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*.
- Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE: Life Sciences Education, (15)*. Hentet fra <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5132390/>
- Bouffard-Bouchard, T. (1990). Influence of Self-efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*, 353-363.

- Campbell, T. (2013). In-school ability grouping and the month of birth effect - Preliminary evidence from the Millennium Cohort Study. *Working paper 2013/1*.
- Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 79-92.
- Cohen, L., Manion, L. & Morrison, K. (2018). *Research Methods in Education* (8. utg.). London and New York: Routledge.
- Collins, J. (1982). *Self-Efficacy and Ability in Achievement Behavior*. New York: paper presented at the meeting of the American Educational Research Association.
- Cools, S., Schöne, P. & Strøm, M. (2017). Forsyvnings i skolestart: Hvilken rolle spiller kjønn og sosial bakgrunn? . Hentet fra https://www.idunn.no/spa/2017/04/forskyvninger_i_skolestart_hvilken_rolle_spiller_kjoenn_og
- Crawford, C., Dearden, L. & Greaves, E. (2013a). The impact of age within academic year on adult. *IFS Working Papers W13/07*.
- Crawford, C., Dearden, L. & Greaves, E. (2013b). *When You Are Born Matters: Evidence for England: IFS Report R80*. London.
- Doménech-Betoret, F., Abellán-Roselló, L. & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction and academic achievement: The mediator role of students' expectancy-value beliefs. *Frontiers in Psychology, 8*. <https://doi.org/doi:10.3389/fpsyg.2017.01193>
- Engelhard Jr, G. (2013). *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*. New York and London: Routledge.
- Foldnes, N., Grønneberg, S. & Hermansen, G. H. (2018). *Statistikk og dataanalyse - En moderne innføring*. Oslo: Cappelen Damm.
- Foyn, T. (2019). A call for nuancing the debate on gender, education and mathematics in Norway. *Eleventh Congress of the European Society for Research in Mathematics Education*.
- Friborg, O. (2010). Klassisk testteori og utvikling av spørreinstrumenter. I M. Martinussen (Red.), *Kvantitativ forskningsmetodologi i samfunns- og helsefag*. Bergen: Fagbokforlaget.
- Frøseth, M. W., Hovdhaugen, E., Høst, H. & Vibe, N. (2008). *Tilbudsstruktur og gjennomføring i videregående opplæring* (40/2008). Norsk institutt for studier av innovasjon, forskning og utdanning. Hentet fra <https://nifu.brage.unit.no/nifu-xmloi/bitstream/handle/11250/284580/NIFUrapport2008-40.pdf?sequence=1&isAllowed=y>
- Gabrielsen, E. & Lundetræ, K. (2017). Indikerer de norske PIRLS-resultatene et behov for å justere retningslinjene for skolestartalder? I E. Gabrielsen (Red.), *Klar framgang! Lesferdigheter på 4. og 5. trinn i et femtenårsperspektiv*. Universitetsforlaget.
- Gallagher, A. M. & Kaufman, J. C. (2005). Gender differences in mathematics: What we know and what we need to know. I A. M. Gallagher & J. C. Kaufman (Red.), *Gender differences in mathematics*. New York: Cambridge University Press.
- Gledhill, J., Ford, T. & Goodman, R. (2002). Does season of birth matter? The relationship between age within the school year (season of birth) and educational difficulties among a representative general population sample of children and adolescents (age 5-15) in Great Britain. *Research in Education 68*, 41-47.
- Glen, S. (2016). Rasch Model/ Rasch Analysis: Definition, Examples. Hentet fra <https://www.statisticshowto.com/rasch-model/>
- Granger, C. V. (2008). Rasch Analysis is Important to Understand and use for Measurement. Hentet fra <https://www.rasch.org/rmt/rmt213d.htm>

- Grønmo, L. S., Lindquist, M., Arora, A. & Mullis, I. V. S. (2013). TIMS 2015 Mathematics Framework. I I. V. S. Mullis & M. O. Martin (Red.), *TIMSS 2015 Assessment Frameworks*. Chestnut Hills: TIMSS & PIRLS International Study Center.
- Grønmo, S. (2020). bias i forskning. Hentet fra https://snl.no/bias_i_forskning
- Hackett, G. & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education*.
- Halvorsen, K. (2008). *Å forske på samfunnet: En innføring i samfunnsvitenskapelig metode* (5. utg.). Oslo: Cappelen Akademisk Forlag.
- IEA. (u.d). TIMSS - Trends in International Mathematics and Science Study. Hentet fra iea.nl/studies/iea/timss
- Jensen, F. & Nordtvedt, G. A. (2013). Holdninger til matematikk. I M. Kjærnsli & R. V. Olsen (Red.), *Fortsatt en vei å gå - Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012* (s. 97-117). Oslo: Universitetsforlaget.
- Jensen, F., Pettersen, A., Frønes, T. S., Kjærnsli, M., Rohatgi, A., Eriksen, A. & Narvhus, E. K. (2019). *PISA 2018. Norske elevers kompetanse i lesning, matematikk og naturfag*. Oslo: Universitetsforlaget. Hentet fra https://www.uv.uio.no/ils/forskning/prosjekter/pisa/publikasjoner/publikasjoner/pisa2018_kortrapport.pdf
- Johannessen, A., Tuft, P. A. & Christoffersen, L. (2012). *Introduksjon til samfunnsvitenskapelig metode*. Oslo: Abstrakt forlag.
- Jussim, L., Robustelli, S. L. & Cain, T. R. (2009). Teacher expectations and self-fulfilling prophecies. I K. R. Wentzel & A. Wigfield (Red.), *Handbook of Motivation at School* New York: Routledge.
- Karlstad, Ø., Furu, K., Stoltenberg, C., Håberg, S. E. & Bakken, I. J. (2017). ADHD treatment and diagnosis in relation to children's birth month: Nationwide cohort study from Norway. *Scandinavian Journal of Public Health*, 45, 343-349.
- Kleven, T. A. & Hjørdemaal, F. R. (2018). *Innføring i pedagogisk forskningsmetode: En hjelp til kritisk tolkning og vurdering* (3. utg.). Bergen: Fagbokforlaget.
- Kunnskapsdepartementet. (2011). Fra matteskrekk til mattemestring. Hentet fra https://www.regjeringen.no/globalassets/upload/kd/vedlegg/grunnskole/strategiplaner/matematikk_aug_2011.pdf
- Kunnskapsdepartementet. (2020). *Overordnet del - Undervisning og tilpasset opplæring*. Hentet fra <https://www.udir.no/lk20/overordnet-del/3.-prinsipper-for-skolens-praksis/3.2-undervisning-og-tilpasset-opplaring/>
- Kaarstein, H. & Nilsen, T. (2016). Motivasjon. I O. K. Bergem, H. Kaarstein & T. Nilsen (Red.), *Vi kan lykkes i realfag - Resultater og analyser fra TIMSS 2015*. Oslo: Universitetsforlaget.
- Kaarstein, H., Radišić, J., Lehre, A.-C., Nilsen, T. & Bergem, O. K. (2020). *TIMSS 2019 - Kortrapport*. Universitetet i Oslo.
- Lien, L., Tambs, K., Oppedal, B., Heyerdahl, S. & Bjertness, E. (2005). Is relative young age within a school year a risk factor for mental health problems and poor school performance? A population-based cross-sectional study of adolescents in Oslo, Norway. *BMC Public Health*, 5:102.
- Linacre, J. M. (2002). Differential Item Functioning and Differential Test Functioning (DIF & DTF). Hentet fra <https://www.rasch.org/rmt/rmt163g.htm>
- Linacre, J. M. (2012). Windsteps Tutorial 4. Hentet fra https://www.winsteps.com/a/winsteps-tutorial-4.pdf?fbclid=IwAR2IW8GOgNLTs7Dt9Cxae_FDy7vc1ZOCUOhX5WR5WuYXs0IY125WPWO0Dic

- Linacre, J. M. (u.d.). Fit diagnosis: infit outfit mean-square standardized. Hentet fra <https://www.winsteps.com/winman/misfitdiagnosis.htm>
- Martin, R. P., Foels, P., Clanton, G. & Moon, K. (2004). Season of birth is related to child retention rates, achievement, and rate of diagnosis of specific LD. *Journal of learning disabilities*, 37, 307-317.
- Nordtvedt, G. A. (2013). Resultater i matematikk. I M. Kjærnsli & R. V. Olsen (Red.), *Fortsatt en vei å gå - Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012*. Oslo: Universitetsforlaget.
- NOU 2019: 3. (2019). *Nye sjanser - Bedre læring: Kjønnforskjeller i skoleprestasjoner og utdanningsløp*. Hentet fra <https://www.regjeringen.no/contentassets/8b06e9565c9e403497cc79b9fdf5e177/no/pdfs/nou201920190003000dddpdfs.pdf>
- Olsen, R. V. & Björnsson, J. K. (2018). Fødselsmåned og skoleprestasjoner. I R. V. Olsen & J. K. Björnsson (Red.), *Tjue år med TIMSS og PISA i Norge*. Oslo: Universitetsforlaget.
- Olsen, R. V. & Blömeke, S. (2018). Hva forklarer endringer i elevenes matematikkprestasjon over tid? . I J. K. Björnsson & R. V. Olsen (Red.), *Tjue år med TIMSS og PISA i Norge - Trender og analyser*. Oslo: Universitetsforlaget.
- Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. I A. M. Gallagher & J. C. Kaufman (Red.), *Gender differences in mathematics - An integrative psychological approach*. New York: Cambridge University Press.
- Pajares, F. (2006). Self-efficacy during childhood and adolescence: Implications for teacher and parents. I F. Pajares & T. Urdan (Red.), *Self-efficacy beliefs of adolescents*. Greenwich, Connecticut: IAP.
- Pajares, F. & Graham, L. (1999). Self-Efficacy, Motivation Constructs, and Mathematics Performance of Entering Middle School Students. *Contemporary Educational Psychology*, 24, 124-139.
- Pajares, F. & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, 20, 426-443.
- Pajares, F. & Miller, D. M. (1995). Mathematics self-efficacy and mathematics performance: The need of specificity of assessment. *Journal of Counseling Psychology*, 42, 190-198.
- Pampaka, M., Williams, J., Hutcheson, G., Wake, G., Black, L., Davis, P. & Hernandez-Martinez, P. (2012). The association between mathematics pedagogy and learners' dispositions for university study. *British Educational Research Journal*, 38, 473-496. <https://doi.org/10.1080/01411926.2011.555518>
- Pietsch, J., Walker, R. & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of educational psychology*, 95, 589-603.
- Randhawa, B. S., Beamer, J. E. & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of educational psychology*, 85, 41-48.
- Reed, K. E., Parry, D. A. & Sandercock, G. R. M. (2017). Maturational and social factors contributing to relative age effects in school sports: Data from the London Youth Games. *The Scandinavian Journal of Medicine & Science in Sports*, 27, 2070-2079. <https://doi.org/10.1111/sms.12815>
- Ringdal, K. (2013). *Enhet og mangfold: Samfunnsvitenskaplig forskning og kvantitativ metode* (3. utg.). Bergen: Fagbokforlaget.
- Roick, J. & Ringeisen, T. (2018). Students' math performance in higher education: Examining the role of self-regulated learning and self-efficacy. *Learning and*

- Individual Differences*, 65, 148-158. Hentet fra <https://doi.org/10.1016/j.lindif.2018.05.018>
- Rosenlund, M. R. & Gulaker, D. T. F. (2018). Hvordan skape motivasjon for matematikk? I T. A. Fiskum, D. Gulaker & H. P. Andersen (Red.), *Den engasjerte eleven - Undrende, utforskende og aktiviserende undervisning i skolen* (s. 169-187). Oslo: Cappelen Damm Akademiske.
- Sharp, C., Hutchison, D. & Whetton, C. (1994). How do season of birth and length of schooling affect children's attainment at key stage 1? *Educational Research*, 36:2, 107-121.
- Skaalvik, E. M., Federici, R. A. & Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research*, 72, 129-136.
- Skaalvik, E. M. & Skaalvik, S. (2018). *Skolen som læringsarena - Selvoppfatning, motivasjon og læring*. Oslo: Universitetsforlaget.
- Solli, I. F. (2017). Left behind by birth month. *Education Economics*, 25:4, 323-346. <https://doi.org/10.1080/09645292.2017.1287881>
- Statistisk sentralbyrå. (2017). Guttene havner bakpå. Hentet fra <https://www.ssb.no/utdanning/artikler-og-publikasjoner/guttene-havner-bakpa>
- Statistisk sentralbyrå. (2020, 17. desember). Elevar i grunnskolen. Hentet fra <https://www.ssb.no/utdanning/statistikker/utgrs/aar>
- Street, K. E. S., Malmberg, L.-E. & Stylianides, G. J. (2017). Level, strengt, and facet-specific self-efficacy in mathematics test performance. *ZDM Mathematics Education*, 49, 279-395. Hentet fra <https://doi.org/10.1007/s11858-017-0833-0>
- Strøm, B. (2004). Student achievement and birthday effects. *Norwegian University of Science and Technology working paper*.
- Tennant, A. & Pallant, J. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. Hentet fra <https://www.rasch.org/rmt/rmt204d.htm>
- Tufte, P. A. (2011). Kvantitativ metode. I K. Fangen & A.-M. Sællerberg (Red.), *Mange ulike metoder*. Oslo: Gyldendal Akademisk.
- Usher, E. L. & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78, 751-796.
- Usher, E. L. & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34, 89-101.
- Utdanningsdirektoratet. (2016). TIMSS. Hentet fra <https://www.udir.no/tall-og-forskning/internasjonale-studier/timss/>
- Utdanningsdirektoratet. (2018). Metodegrunnlaget for nasjonale prøver. Hentet fra <https://www.udir.no/globalassets/filer/vurdering/nasjonaleprover/metodegrunnlag-for-nasjonale-prover-august-2018.pdf>
- Vibe, N., Frøseth, M. W., Hovdhaugen, E. & Markussen, E. (2012). *Strukturer og konjunkturer - Evaluering av Kunnskapsløftet: Sluttrapport av prosjektet «Tilbudsstruktur, gjennomføring og kompetanseoppnåelse i videregående opplæring»* (26/2012). Oslo: Nordisk institutt for studier av innovasjon, forskning og utdanning. Hentet fra https://www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2012/nifu_struktur.pdf
- Wattie, N., Coble, S. & Baker, J. (2008). Towards a unified understanding of relativ age effects. *Journal of Sport Sciences*, 26(13), 1403-1409.
- Wolfe, E. W. & Smith Jr, E. V. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I - instrument development tools. *Journal of Applied Measurement*, 8, 97-123.

- Wolfe, E. W. & Smith Jr, E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: part II - validation activities. *Journal of Applied Measurement*, 8, 204-234.
- Wæge, K. & Nostrati, M. (2018). *Motivasjon i matematikk*. Oslo: Universitetsforlaget.
- Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, 8, 187-203.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. I A. Bandura (Red.), *Self-efficacy in changing societies*. New York: Cambridge University Press.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82-91.
- Zimmerman, B. J. & Clearly, T. J. (2006). Adolescents' development of personal agency: The role of self-efficacy and self-regulatory skills. I F. Pajares & T. Urdan (Red.), *Self-efficacy beliefs of adolescents*. Greenwich, Connecticut: IAP.

Vedlegg 1 - Hovedundersøkelsen

DEL 3 er endret fra slik de opprinnelig var i hovedundersøkelsen, slik at oppgavenummeret samsvarer med mestringsforventningsnummereringen som er brukt i analysene.



- Kjønn** Jente Gutt
- Fødselsår** 2006 Annet
- Morsmål** Norsk Annet

Fødselsmåned

- Januar Februar Mars April
- Mai Juni Juli August
- September Oktober November Desember

DEL 1

Hvor trygg er du på om du greier å løse ulike matteoppgaver?

Svar på en skala fra 1-6, hvor 1 betyr «*veldig utrygg*» og 6 betyr «*veldig trygg*».

	1	2	3	4	5	6
Oppgave 1						
Oppgave 2						
Oppgave 3						
Oppgave 4						
Oppgave 5						
Oppgave 6						
Oppgave 7						
Oppgave 8						
Oppgave 9						
Oppgave 10						
Oppgave 11						

Oppgave 12						
------------	--	--	--	--	--	--

DEL 2

Hvor godt stemmer påstandene for deg? Gi svar på en skala fra 1 – 6.

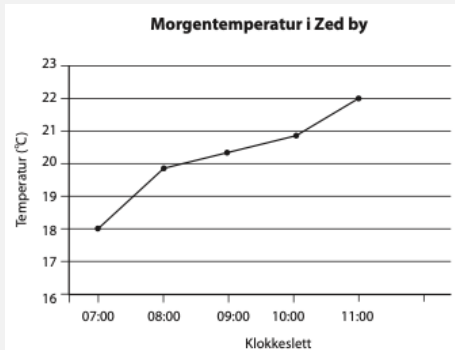
1 betyr «stemmer ikke i det hele tatt» og 6 betyr «stemmer veldig godt».

	1	2	3	4	5	6
1. Jeg får gode karakterer på matteprøver.						
2. Jeg har alltid lykkes i matte.						
3. Selv når jeg jobber veldig hardt, gjør jeg det dårlig i matte.						
4. Jeg fikk god karakter i matte til sommeren i 8.klasse.						
5. Jeg er flink til å løse ulike matteoppgaver.						
6. Jeg får til å løse selv de vanskeligste oppgavene i matte.						
7. Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre.						
8. Når jeg ser mattelæreren min løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
9. Å se andre elever gjøre det bedre enn meg i matematikk, pusher meg til å gjøre det bedre.						
10. Når jeg ser en medelev løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
11. Jeg kan se for meg at jeg jobber med vanskelige matteoppgaver og lykkes med å løse dem.						
12. Jeg konkurrerer med meg selv i matte.						
13. Mattelærerne mine har fortalt meg at jeg er flink i matte.						
14. Andre har fortalt meg at jeg er veldig flink i matte.						
15. Voksne i familien min har fortalt meg at jeg er flink i matte.						
16. Jeg har fått mye skryt for mine evner i matte.						
17. Medelevene mine har fortalt meg at jeg er flink i matte.						
18. Klassekameratene mine liker å jobbe med meg i matte, fordi de syns jeg er god.						
19. Jeg føler meg stresset og nervøs bare ved å være i mattetimene.						
20. Å jobbe med matte tapper meg for energi.						
21. Jeg blir stresset med en gang jeg begynner å jobbe med matte.						
22. Jeg får jernteppe og klarer ikke å tenke klart når jeg gjør matteoppgaver						
23. Jeg blir deppa av tanken på å gjøre matte.						
24. Jeg blir anspent i hele kroppen når jeg må jobbe med matte.						

DEL 3

Regn ut oppgavene nedenfor.

Oppgave 4



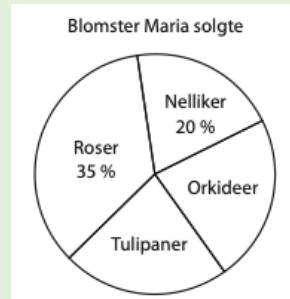
Grafen

viser temperaturen hver time fra klokka 07:00 til klokka 11:00.

Anslå temperaturen klokka 09.30.

Svar: _____ °C

Oppgave 7



Maria solgte 4 typer blomster. Hun solgte like mange tulipaner og orkideer.

Hvor mange prosent av blomstene som ble solgt, var tulipaner?

Svar: _____ %

Oppgave 11

Hva er størst, $\frac{7}{12}$ eller $\frac{2}{3}$?

Svar: _____

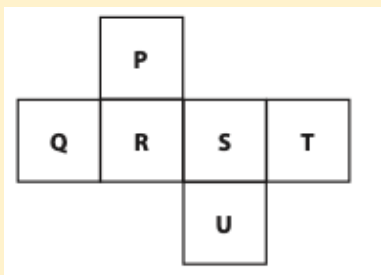
Oppgave 1

Skriv dette som et desimaltall.

$$8 + 50 + \frac{3}{100} + \frac{1}{10}$$

Svar: _____

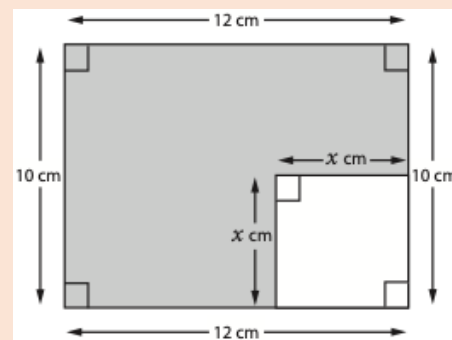
Oppgave 3



Lisa har bygd en terning ved å brette figuren over. Hvilken sideflate står på motsatt side av Q?

Svar: _____

Oppgave 12



Skriv et uttrykk (ved hjelp av x) for arealet av den **gråfargede** delen av figuren.

Svar: _____ cm^2

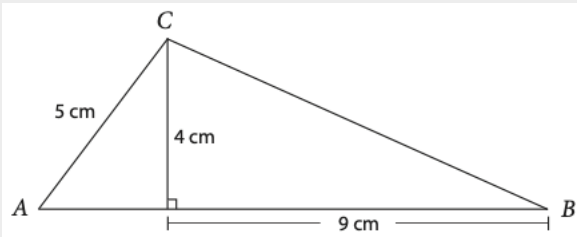
Oppgave 10

En maskin har 7 forskjellige farger på tyggegummiballer. Liv la merke til at folk kjøpte 306 tyggegummiballer, og at 23 av de var blå.

Hva er sannsynligheten for at den neste tyggegummiballen vil være blå?

Svar: _____

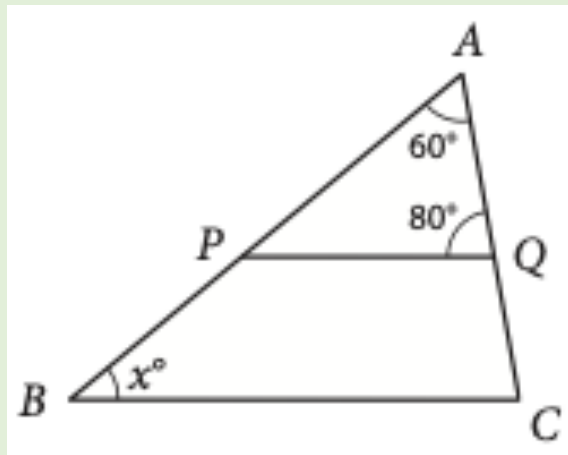
Oppgave 6



Hva er arealet til trekanten ABC?

Svar: _____ cm^2

Oppgave 2



Linjene PQ og BC er parallelle.
Hvor stor er x?

Svar: _____

Oppgave 9

Elevenes favorittprogram på TV



Sektordiagrammet viser hvilken type TV-program 240 elever likte best. Hvor mange elever likte programmer om historie best?

Svar: _____

Oppgave 5

Finn verdiene av x og y slik at begge likningene er sanne.

$$3x + y = 13$$

$$5x - y = 27$$

Svar: $x =$ _____ $y =$ _____

Oppgave 8

$$a = 5 \text{ og } b = 2$$

Hva er verdien av $a^2b - 3(a - b)$?

Svar: _____

Vedlegg 2 – Retningslinjer for gjennomføringen

Presentere hvem vi er, og hvorfor vi er kommet til deres klasse.

Vi presenterer undersøkelsen:

- Undersøkelsen tar ca. en time. Vi setter pris på at du svarer så godt du kan. Påpeke at vi skal ikke teste hvor mye matematikk elevene kan, så ingen trenger å få nerver.
- Undersøkelsen er helt anonym, så ikke skriv navnet ditt på noen av arkene du får utdelt. Det er frivillig å delta, og du kan trekke deg ved å si ifra til meg eller ved å ikke levere inn undersøkelsesheftet.
- Heftet består av tre deler. Vi går gjennom de ulike delene, slik at de vet hva som kommer. Si at dette er et individuelt arbeid, så de kan ikke samarbeide med andre.
Vi deler ut heftene, ber elevene finne frem en blyant og fylle ut forsiden.

Del 1

- Vi tar opp presentasjonen på tavla og går igjennom et eksempel. Her presiserer vi at elevene ikke må krysse mellom rutene, men at de må velge en av rutene. Gå gjennom eksempel, presiser at dette er samme oppsett som resten av DEL 1. Presiser at elevene ikke skal regne ut, men krysse av for trygg de er på at de kan løse oppgaven. Hver oppgave blir lest høyt før vi går rett til hvitt ark og elevene får tid til å krysse. Når vi er ferdig med del 1 presiseres det at det ikke er lov til å bla tilbake i heftet for å endre svarene sine.

Del 2

- Vi ber elevene bla om og informerer elevene om at de skal sette ett kryss pr. påstand. Her er det også viktig at elevene ikke krysser mellom rutene. Må også informere om at det her er en "ny skala" der 1 er "stemmer ikke i det hele tatt" og 6 er "stemmer veldig godt".
Gi beskjed om at elevene kan fortsette på del 3 når de er ferdig, og begynne å regne oppgavene. Informere om at de får et kladdark, der de kan gjøre utregninger hvis de trenger, og fylle inn svaret i heftet. Rekke opp hånda når man er ferdig med å svare på alt.

Spørsmål vi fikk i pilot-undersøkelsen

- "Hva regnes som gode karakterer?" - vi må presisere at det er elevenes egen mening som gjelder.
- Hva er morsmål - det språket som blir snakket hjemme
- Spørsmål om hjelp til utregning – vi kan ikke hjelpe elevene.

Vi må sjekke at forsiden er utfylt når vi samler inn undersøkelsen.

Vedlegg 3 – Informasjonsskriv til foresatte

Informasjon om gjennomføring av spørreundersøkelse i uke 43/44

Hei!

Mitt navn er Anna Iversen og jeg er lærerstudent ved OsloMet. I forbindelse med min mastergrad, er jeg med i forskningsprosjektet SESAM – *Self-efficacy, its sources and achievement in mathematics*. Nå reiser jeg rundt til ulike klasser på 9.trinn for å gjennomføre en undersøkelse i matematikk.

Undersøkelsen består av tre deler, og innebærer både avkrysningsoppgaver og matematikkoppgaver som skal løses.

Undersøkelsen er helt anonym, hverken skole, klasse eller elev er mulig å spore.

Undersøkelsen tar omtrent 60 minutter å gjennomføre, og det er frivillig for elevene å delta i undersøkelsen.

Jeg er veldig glad for å få komme til deres skole!

Hilsen Anna Iversen



Vedlegg 4 – Pilotundersøkelsen

Hei!

Denne undersøkelsen gjennomføres av forskningsgruppa SESAM fra OsloMet. Vi er tre masterstudenter, som sammen med vår veileder, samler inn data på 9. trinn i Norge.

Vi setter stor pris på at du svarer så godt du kan. Det er frivillig å delta i undersøkelsen. Svaret ditt er anonymt, så du skal ikke skrive navnet ditt på arket. Det er kun forskere i SESAM som skal se besvarelsen din.

Undersøkelsen består av tre deler – noe avkrysning og noe regning.



Kryss av for det alternativet som stemmer for deg:

Kjønn

- Jente Gutt

Fødselsår

- 2006 Annet

Fødselsmåned

- Januar Februar Mars April
 Mai Juni Juli August
 September Oktober November Desember

Morsmål

- Norsk Annet

DEL 1

Hvor trygg er du på om du greier å løse ulike matteoppgaver?

Svar på en skala fra 1-6, hvor 1 betyr «*veldig utrygg*» og 6 betyr «*veldig trygg*».

Oppgavene blir presentert på tavla og lest høyt.

	<i>Veldig utrygg</i>					<i>Veldig trygg</i>
	1	2	3	4	5	6
Oppgave 1						
Oppgave 2						
Oppgave 3						
Oppgave 4						
Oppgave 5						
Oppgave 6						
Oppgave 7						
Oppgave 8						
Oppgave 9						
Oppgave 10						
Oppgave 11						
Oppgave 12						
Oppgave 13						
Oppgave 14						
Oppgave 15						
Oppgave 16						

DEL 2

Hvor godt stemmer påstandene for deg? Gi svar på en skala fra 1-6, der 1 betyr «stemmer ikke i det hele tatt» og 6 betyr «stemmer veldig godt».

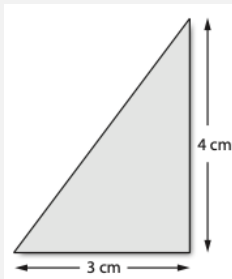
	1	2	3	4	5	6
Jeg får gode karakterer på matteprøver						
Jeg har alltid lykkes i matte						
Selv når jeg jobber veldig hardt, gjør jeg det dårlig i matte						
Jeg fikk god karakter i matte til sommeren i 8.klasse						
Jeg er flink til å løse ulike matteoppgaver						
Jeg får til å løse selv de vanskeligste oppgavene i matte						
Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre.						
Når jeg ser mattelæreren min løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
Å se andre elever gjøre det bedre enn meg i matematikk, pusher meg til å gjøre det bedre.						
Når jeg ser en medelev løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
Jeg kan se for meg at jeg jobber med vanskelige matteoppgaver og lykkes med å løse dem.						
Jeg konkurrerer med meg selv i matte.						
Mattelærerne mine har fortalt meg at jeg er flink i matte.						
Andre har fortalt meg at jeg er veldig flink i matte.						
Voksne i familien min har fortalt meg at jeg er flink i matte.						
Jeg har fått mye skryt for mine evner i matte.						
Medelevene mine har fortalt meg at jeg er flink i matte.						
Klassekameratene mine liker å jobbe med meg i matte, fordi de synes jeg er god.						
Jeg føler meg stresset og nervøs bare ved å være i mattetimene.						
Å jobbe med matte tapper meg for energi.						
Jeg blir stresset med en gang jeg begynner å jobbe med matte.						
Jeg får jernteppe og klarer ikke å tenke klart når jeg gjør matteoppgaver						
Jeg blir deppa av tanken på å gjøre matte.						
Jeg blir anspent i hele kroppen når jeg må jobbe med matte.						

DEL 3

Regn ut oppgavene nedenfor.

Oppgave 1

Andreas har et rektangulært papir.
Han deler det i to og står igjen med denne papirtrekanten.



Beregn lengden av den tredje siden av trekanten.

Svar: _____ cm

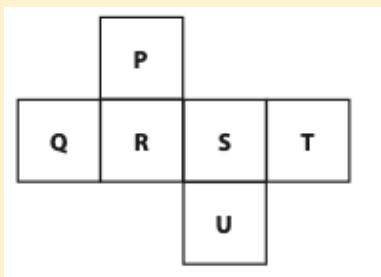
Oppgave 2

$$a = 1 + x \text{ og } b = 1 - x$$

Hva er $a - b$?

Svar: _____

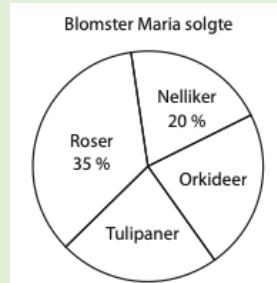
Oppgave 3



Lisa har bygd en terning ved å brette figuren over. Hvilken sideflate står på motsatt side av Q?

Svar: _____

Oppgave 4



Maria solgte 4 typer blomster. Hun solgte like mange tulipaner og orkideer.

Hvor mange prosent av blomstene som ble solgt, var tulipaner?

Svar: _____ %

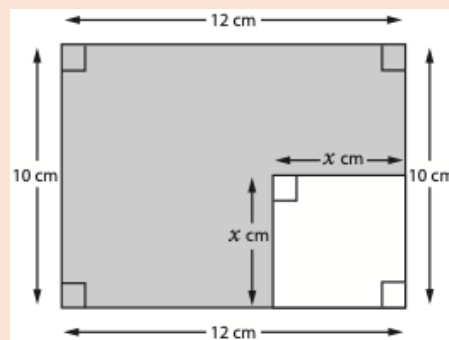
Oppgave 5

Skriv dette som et desimaltall.

$$8 + 50 + \frac{3}{100} + \frac{1}{10}$$

Svar: _____

Oppgave 6



Skriv et uttrykk (ved hjelp av x) for arealet av den gråfargede delen av figuren.

Svar: _____ cm^2

Oppgave 7

En maskin har 7 forskjellige farger på tyggegummiballer. Liv la merke til at folk kjøpte 306 tyggegummiballer, og at 23 av de var blå.

Hva er sannsynligheten for at den neste tyggegummiballen vil være blå?

Svar: _____

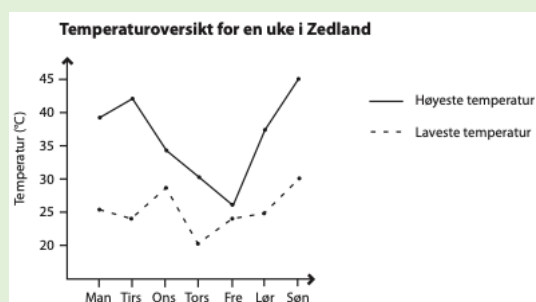
Oppgave 10

Du har en kakeoppskrift som krever 2 egg og 0,3 liter melk. Du vil lage en så stor kake som mulig, og du har 5 egg.

Hvor mange liter melk trenger du da?

Svar: _____ liter

Oppgave 8



Grafene over viser høyeste og laveste temperatur hver dag i løpet av en uke i Zedland. Hvilken dag var forskjellen mellom den høyeste og den laveste temperaturen 10°C ?

Svar: _____

Oppgave 11



Sektordiagrammet viser hvilken type TV-program 240 elever likte best. Hvor mange elever likte programmer om historie best?

Svar: _____

Oppgave 9

Finn verdiene av x og y slik at begge likningene er sanne.

$$3x + y = 13$$

$$5x - y = 27$$

Svar: $x =$ _____ $y =$ _____

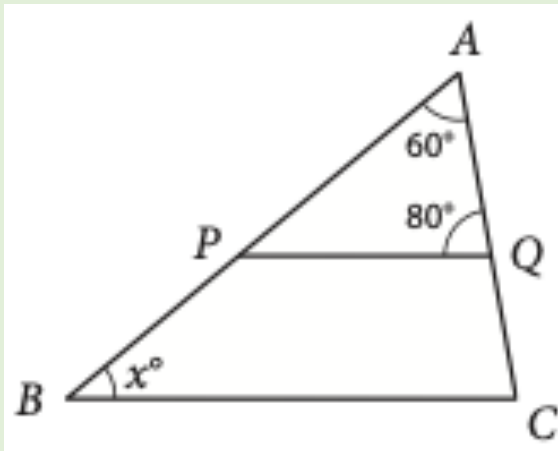
Oppgave 12

$a = 5$ og $b = 2$

Hva er verdien av $a^2b - 3(a - b)$

Svar: _____

Oppgave 13

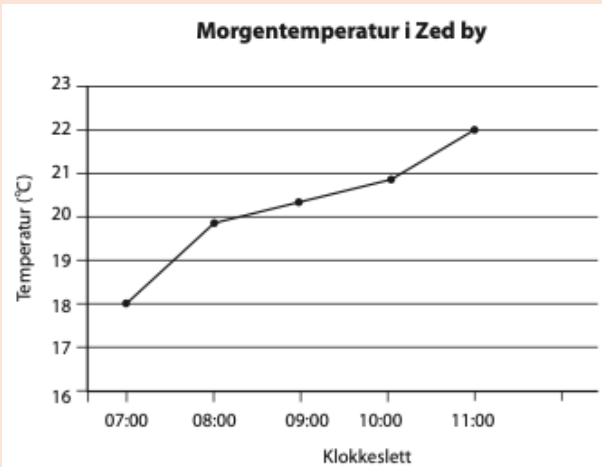


Linjene PQ og BC er parallelle.

Hvor stor er x ?

Svar: _____

Oppgave 15



Grafen viser temperaturen hver time fra klokka 07:00 til klokka 11:00.

Anslå temperaturen klokka 09.30.

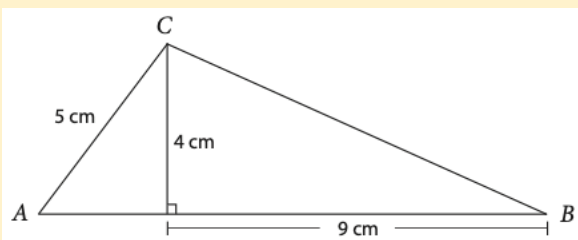
Svar: _____ °C

Oppgave 14

Hva er størst, $\frac{7}{12}$ eller $\frac{2}{3}$?

Svar: _____

Oppgave 16



Hva er arealet til trekanten ABC?

Svar: _____

Vedlegg 5 – Tilbakemelding fra NSD

3.5.2021

Meldeskjema for behandling av personopplysninger



NSD sin vurdering

Prosjekttittel

SESAM - Self-efficacy, it's sources and achievement in mathematics

Referansenummer

152185

Registrert

22.09.2020 av Marie Larsen - s313582@oslomet.no

Behandlingsansvarlig institusjon

OsloMet – storbyuniversitetet / Fakultet for lærerutdanning og internasjonale studier / Institutt for grunnskole- og faglærerutdanning

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Annette Hessen Bjerke, annette.hessen@oslomet.no, tlf: 41254388

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Marie Larsen, s313582@oslomet.no, tlf: 99362324

Prosjektperiode

24.08.2020 - 15.06.2021

Status

30.09.2020 - Vurdert anonym

Vurdering (1)

30.09.2020 - Vurdert anonym

Det er vår vurdering at det ikke skal behandles direkte eller indirekte opplysninger som kan identifisere enkeltpersoner i dette prosjektet, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet den 30.09.2020 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD.

Prosjektet trenger derfor ikke en vurdering fra NSD.

HVA MÅ DU GJØRE DERSOM DU LIKEVEL SKAL BEHANDLE PERSONOPPLYSNINGER?

Dersom prosjektopplegget endres og det likevel blir aktuelt å behandle personopplysninger må du melde dette til NSD ved å oppdatere meldeskjemaet. Vent på svar før du setter i gang med behandlingen av

3.5.2021

Meldeskjema for behandling av personopplysninger

personopplysninger.

VI AVSLUTTER OPPFØLGING AV PROSJEKTET

Siden prosjektet ikke behandler personopplysninger avslutter vi all videre oppfølging.

Lykke til med prosjektet!

Kontaktperson hos NSD: Kajsa Amundsen
Tlf. Personverntjenester: 55 58 21 17 (tast 1)